Journal of Experimental Psychology:
Human Perception and Performance
1991, Vol. 17, No. 1, 3-27

0096-1523/91/$3.00

Tests of an Exemplar Model for Relating Perceptual Classification and Recognition Memory

Robert M. Nosofsky
Indiana University

Experiments were conducted in which Ss made classification, recognition, and similarity judgments for 34 schematic faces. A multidimensional scaling (MDS) solution for the faces was derived on the basis of the similarity judgments. This MDS solution was then used in conjunction with an exemplar–similarity model to accurately predict Ss' classification and recognition judgments. Evidence was provided that Ss allocated attention to the psychological dimensions differentially for classification and recognition. The distribution of attention came close to the ideal-observer distribution for classification, and some tendencies in that direction were observed for recognition. Evidence was also provided for interactive effects of individual exemplar frequencies and similarities on classification and recognition, in accord with the predictions of the exemplar model. Unexpectedly, however, the frequency effects appeared to be larger for classification than for recognition.

The purpose of this study was to provide tests of a model for relating perceptual classification performance and old–new recognition memory. The model under investigation is the *context theory of classification* proposed by Medin and Schaffer (1978) and elaborated by Estes (1986a) and Nosofsky (1984, 1986). According to the context theory, people represent categories by storing individual exemplars in memory, and make classification decisions on the basis of similarity comparisons with the stored exemplars. This exemplar view of category representation strongly motivates the study of relations between classification and recognition, because if individual exemplars are stored in memory during classification learning, this fact should be corroborated by postacquisition recognition memory tests.

Nosofsky (1988a) demonstrated preliminary support for an exemplar-based approach to relating classification and recognition. The approach assumes that classification decisions are based on the similarity of an item to the exemplars of a target category relative to exemplars of contrast categories. Recognition decisions are based on the absolute summed similarity of an item to all exemplars of all categories. This absolute summed similarity gives a measure of overall familiarity, with higher familiarity values leading to higher recognition probabilities. This idea that recognition judgments may be based on a form of summed similarity or a "global match" to information stored in memory serves as a core assumption for a variety of extant theories (e.g., Gillund & Shiffrin, 1984; Hintzman, 1986, 1988; Metcalfe-Eich, 1982; Murdock, 1982; Ratcliff, 1978). Unique contributions of the present work are to demonstrate support for the summed-similarity rule in situations involving perceptual classification learning, and to demonstrate that fine-grained differences in recognition probabilities can be predicted on the basis of fine-grained differences in similarities among items.

A key assumption in the present theory is that a common representational substrate underlies classification and recognition judgments (namely, memories for individual exemplars), but different decision rules govern performance in the two tasks: the relative-similarity rule for classification and the absolute summed-similarity rule for recognition. The assumption of a common representational substrate allows for theoretical parsimony and a unified framework for understanding classification and recognition. The assumption of different decision rules enables the model to predict dissociations between classification and recognition, and various other phenomena which previous investigators have cited as evidence against exemplar-only memory models. For example, Nosofsky (1988a) demonstrated that the model is capable of predicting low correlations between classification and recognition, lack of positive contingencies between correct classification and "old" recognition responses, faltering old–new discrimination with increases in category size, high false-alarm rates for category prototypes and foils that are low distortions of the prototype, and dissociations between classification and typicality judgments. Each of these phenomena has been interpreted by previous investigators as providing evidence for abstract forms of category representation, or for the idea that separate memory systems underlie classification and recognition performance (e.g., Anderson, Kline, & Beasley, 1979; Metcalfe & Fisher, 1986; Omohundro, 1981).

Although Nosofsky (1988a) provided quantitative tests of the proposed exemplar framework, a limitation of this earlier work was that fairly gross-level relations between classification and recognition were investigated. For example, Nosofsky (1988a) showed that in the traditional "prototype distortion" paradigms (e.g., Homa, Cross, Cornell, Goldman, & Schwartz, 1973; Posner & Keele, 1970), the exemplar model roughly predicted levels of classification and recognition performance

This work was supported by Grant BNS 87-19938 from the National Science Foundation to Indiana University. I thank Stephen K. Reed for generously providing me with information necessary to replicate and extend his 1972 experiment. I also thank Danny Ennis, Doug Hintzman, and Doug Medin for their helpful criticisms of an earlier version of this article.

Correspondence concerning this article should be addressed to Robert M. Nosofsky, Psychology Department, Indiana University, Bloomington, Indiana 47405.

for the prototypes, old distortions, and low-level and high-level new distortions.

The present research goes beyond this earlier work by studying detailed quantitative relations between classification and recognition performance for individual items having fine-grained differences in similarity to other items in the set. Theoretical analyses are also conducted that investigate whether subjects may modify their attentional weightings of the component dimensions of the stimuli when making classification and recognition decisions. The idea that selective attention modifies similarities between exemplars has served as a cornerstone in allowing the model to characterize relations between *identification*, which is a choice experiment involving a one-to-one stimulus–response mapping, and *classification*, which involves a many-to-one stimulus–response mapping (e.g., Nosofsky, 1984, 1986, 1987; Shepard, Hovland, & Jenkins, 1961). The present research extends this earlier work by investigating the role of selective attention in shaping the classification–recognition relation. Finally, experiments are conducted in the present research that investigate the role of individual item frequency in classification and recognition (cf. Estes, 1986b; Nosofsky, 1988c).

Experiment 1A

Experiment 1A was a baseline condition for studying the classification–recognition relation. The stimuli and category structure that were used were the same as those used in Reed's (1972) seminal study, which was among the first to contrast the quantitative predictions of competing models of classification. I decided to replicate and extend Reed's (1972) experiment, because it is one of few published studies that reports an advantage for the quantitative predictions of a pure prototype model over an exemplar-based model in situations in which categories are learned by way of induction over the exemplars.[1] However, the exemplar model that Reed (1972) tested was an average-distance model, whereas the context model is a summed-similarity model. Because the context model assumes that the relation between similarity and psychological distance is highly nonlinear, its predictions are often dramatically different from those of average-distance and prototype models (e.g., Estes, 1986a; Medin & Schaffer, 1978; Nosofsky, 1984, 1987; Shepard, 1958, 1987). Unfortunately, Reed (1972) reported only the summary fits for the various models he tested; the actual classification data were not reported. Therefore, it was necessary to repeat his experiment to test the quantitative predictions of the context model.

The stimuli used in Reed's (1972) study and the present one were schematic faces varying along four continuous dimensions: eye height, eye separation, nose length, and mouth height. Subjects learned to classify 10 faces into two categories of 5 faces each. The categories of faces are illustrated in Figure 1. Following a training phase, a transfer phase was conducted in which subjects classified each of the 10 old faces plus 24 additional faces formed by new combinations of the values on the component dimensions. The present experiment extended Reed's (1972) study by also collecting old–new recognition judgments for the 34 faces. A preliminary similarity-ratings study was conducted to derive a multidimensional scaling (MDS) solution for the 34 faces. This MDS solution was then used in conjunction with the context model to generate quantitative predictions of subjects' classification and recognition judgments.

Method

Subjects

Subjects were 138 undergraduates from Indiana University who either participated as part of an introductory psychology course requirement or were paid. There were 80 subjects in the classification experiment and 58 subjects in the similarity-ratings study. The data of 10 subjects in the similarity study were not included in the analyses because they correlated lowly (<.50) with the averaged ratings of the group. All subjects were tested individually and participated in only one condition.

Stimuli and Apparatus

The stimuli were 34 schematic faces varying along four continuous dimensions: eye height, eye separation, nose length, and mouth height. Faces 1–32 were constructed from combinations of three possible values along each of the four dimensions. (Note that these faces form only a subset of the total set of 81 stimuli that could have been generated from the four trinary-valued dimensions. Reed [1972] selected the faces, however, such that all possible pairwise combinations of dimension values were represented in the subset.) As explained later, Faces 33 and 34 had intermediate values on each of the four dimensions. The stimuli were generated on an IBM PC and appeared essentially as illustrated in Figure 1. The physical specifications for the 34 faces are provided in Appendix A. Two pairs of stimuli, Faces 11 and 26 and Faces 18 and 30, were identical in Reed's (1972) study, and the same procedure was followed here. Thus, there were 32 unique faces.

Procedure

In the similarity-ratings study, all 496 unique pairs of faces were presented and subjects rated their similarity on a scale from *most dissimilar* (1) to *most similar* (10). The order of presentation of the pairs was randomized for each subject, as was the left–right placement of the faces on the screen. There were 30 practice trials preceding the 496 ratings, with the practice pairs selected randomly for each subject. Subjects were urged to use the full range of similarity ratings.

The classification experiment consisted of a training phase followed by a transfer phase. On each trial of the training phase 1 of the 10 assigned exemplars was presented and a subject judged whether it belonged to Category 1 or 2. Feedback was provided following the response. The training phase was organized into 12 blocks of 10 trials, with each of the 10 faces presented once during each block. Order of presentation of the faces was randomized within each block.

Following the training phase there was a transfer phase in which all 34 faces were presented. The transfer phase was organized into two blocks of 34 trials each, with each face presented once during a block. Order of presentation of the faces was randomized within each block. On each trial, a subject judged whether the face belonged to Category 1 or 2 and then judged whether the face was "old" or "new." Subjects also gave confidence ratings for their recognition judgments,

[1] The term *prototype model* is used in a specific sense in this article to refer to a model in which the category representation is assumed to be a single point corresponding to the central tendency (i.e., the centroid) of all category exemplars in a multidimensional psychological space.

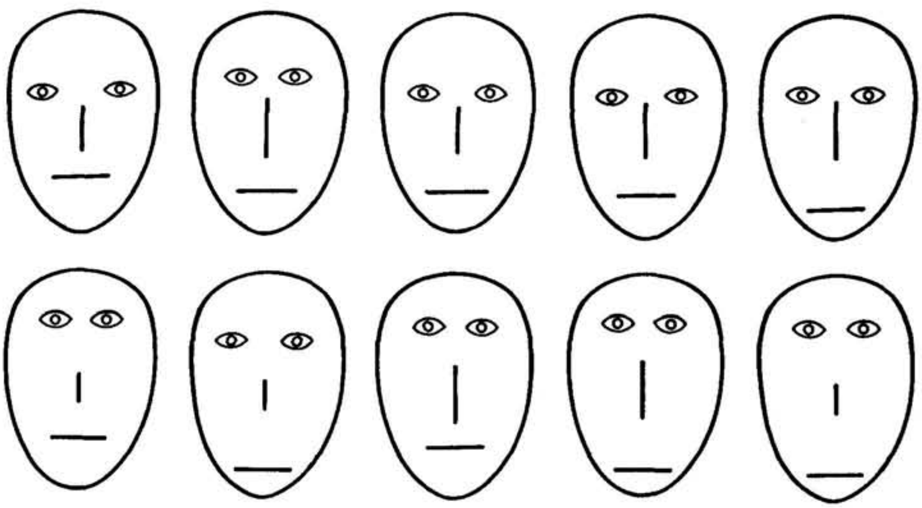

Figure 1. Illustration of the schematic faces used as training items in the present experiments. (Faces in top and bottom rows are members of Category 1 and 2, respectively. *Note.* From "Perceptual vs. conceptual categorization" by S. K. Reed and M. P. Friedman, 1973, *Memory & Cognition, 1,* p. 158. Copyright 1973 by the Psychonomic Society. Reprinted by permission.)

but these ratings are not analyzed in this article. Subjects were instructed to judge as "old" only those faces that had been presented during the training phase. No feedback was presented during the transfer phase, and subjects did not know prior to the start of transfer that their recognition performance would be tested.

Results and Theoretical Analyses

Multidimensional Scaling Analysis

The averaged similarity ratings were used as input to a standard MDS program (KYST; Kruskal, Young, & Seery, 1973). Because of the integral nature of the stimulus dimensions (Lockhead, 1970), a Euclidean metric was specified in the analyses (Garner, 1974; Shepard, 1958, 1987). A four-dimensional solution yielded a good fit to the similarity-ratings data (stress = 0.053), accounting for 94.8% of the linearly explained variance. Higher dimensional solutions led to little improvement in fit and I was unable to interpret the additional dimensions. The coordinates for the MDS solution, after rotation to achieve maximal correspondence with the physically manipulated dimensions of eye height, eye separation, nose length, and mouth height, are reported in Appendix A. These MDS coordinates are used in all subsequent theoretical analyses in this article.

The four-dimensional solution is illustrated in Figure 2. It is evident from inspection that the derived psychological

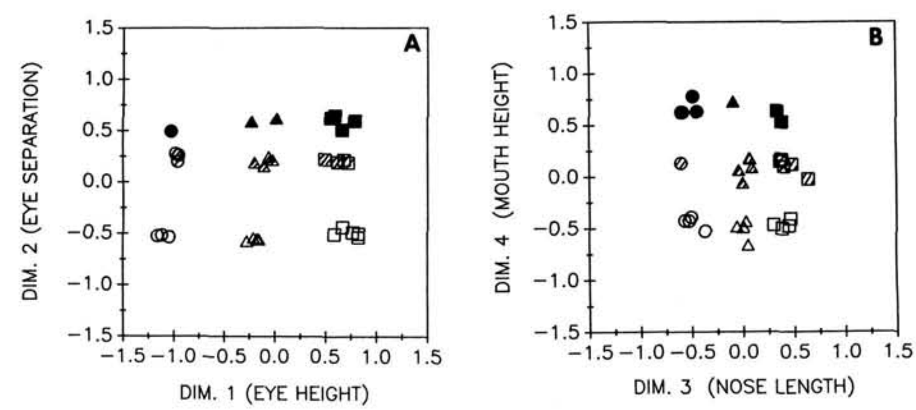

Figure 2. Multidimensional scaling solution for the 34 schematic faces. (Panel A: plot of Dimension 1 [eye height] versus Dimension 2 [eye separation]. Panel B: plot of Dimension 3 [nose length] versus Dimension 4 [mouth height]. Circles, triangles, and squares correspond to physical values 1, 2, and 3, respectively, on Dimensions 1 and 3. Open, cross-hatched, and solid shapes correspond to physical values 1, 2, and 3, respectively, on Dimensions 2 and 4. It is evident from inspection that the stimuli are located systematically such that the underlying psychological dimensions can be interpreted in terms of the variations on the physically manipulated dimensions.)

dimensions are readily interpretable in terms of the physically manipulated dimensions. This result was surprising because I had expected that various configural properties of the faces would be important in influencing subjects' similarity judgments (e.g., Cox & Wallsten, 1987; Pomerantz & Garner, 1973). One configural property of particular interest is the distance between the level of the eyes and the top of the nose, which serves as a perfectly valid cue for discriminating the categories of faces (see Figure 1). Whether this configural property becomes salient in influencing subjects' classification judgments is a question that is addressed in an upcoming section of this article.

The category structure is shown embedded in the MDS solution in Figure 3. As can be seen, four faces in each category form a prototypical cluster in the plot of Dimension 1 versus Dimension 2, whereas the fifth face in each category is rather atypical and falls near the category boundary illustrated in Figure 3. Faces 1–5 correspond to the top five faces in Figure 1, whereas Faces 6–10 correspond to the bottom five faces. The atypical faces, close to the category boundary, are the second faces in each row of Figure 1. As illustrated in Figure 3, the categories are linearly separable (e.g., Medin & Schwanenflugel, 1981; Reed, 1972), which is a necessary requirement for accurate classification through the use of a prototype strategy. Figure 3 also shows the physical prototypes for each category (Faces 33 and 34), which were formed by averaging the physical dimension values associated with the five faces in each category; the psychological prototypes for each category (P1 and P2), defined by averaging the psychological dimension values associated with the five faces in each category; and two control faces (Faces 31 and 32), which are discussed later. The physical prototypes and control faces are actual stimuli that were presented to subjects during the transfer phase, whereas the psychological prototypes are theoretical constructs.

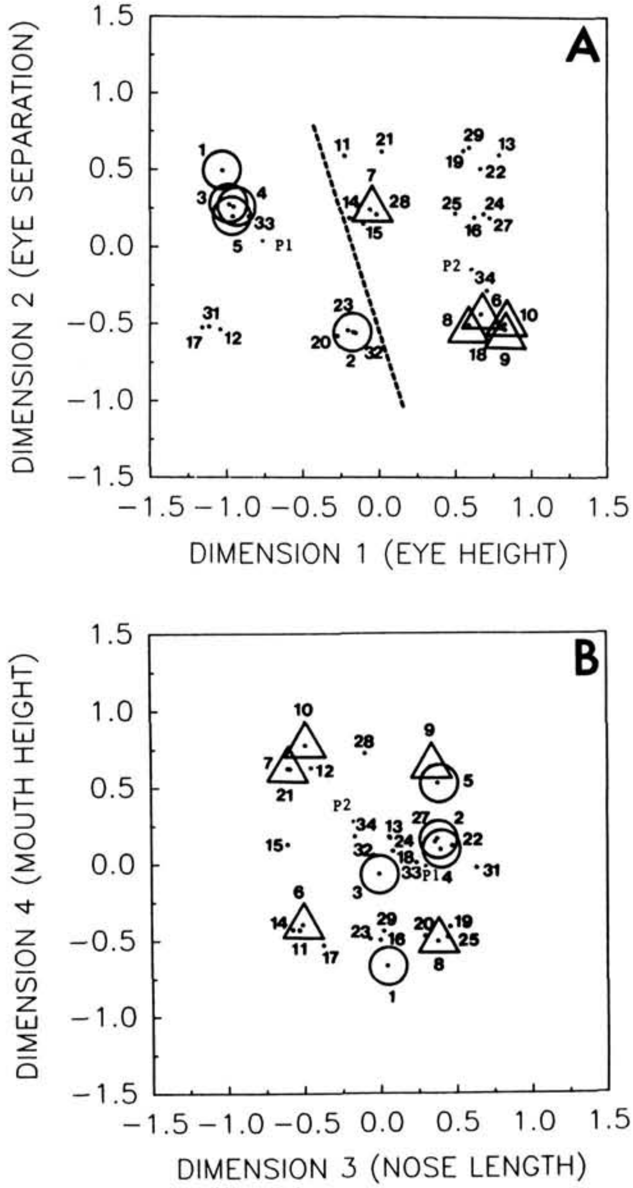

Figure 3. Illustration of the category structures and transfer stimuli embedded in the MDS solution. (Stimuli enclosed by circles denote training exemplars for Category 1, and stimuli enclosed by triangles denote training exemplars for Category 2. The psychological prototypes [P1 and P2] shown were computed by averaging over the psychological coordinate values associated with the stimuli of each category. Stimuli 33 and 34 are the physical prototypes, and Stimuli 31 and 32 are control faces [see text].)

Classification and Recognition

The classification and recognition data obtained during the transfer phase are reported in Table 1. The table shows the probability with which each individual face was classified in Category 1 and the probability with which it was judged as "old." To gain preliminary insight into the classification–recognition relation, Figure 4 plots the observed recognition probabilities against a classification confidence measure. The classification confidence associated with face i [$cc(i)$] is defined as $cc(i) = 2 * |P(R_1|i) - .5|$, where $P(R_1|i)$ is the probability with which face i was classified in Category 1. Thus, for faces that are classified with probability near unity into either Category 1 or 2, the confidence measure will be near unity, whereas for faces that are classified with probability close to .5, the confidence measure will be near zero. The motivation behind plotting the recognition probabilities against the classification confidences is the hypothesis that subjects might judge as "old" those faces that they are fairly certain belong in Category 1 or 2 and judge as "new" those faces that are unclear cases (e.g., Anderson et al., 1979; Metcalfe & Fisher, 1986). As is clear from inspection of the scatterplot, however, there is little correlation between the observed recognition probabilities and the classification confidences ($r = .36$). Indeed, various examples exist in which classification confidence is low, yet recognition probability is high (Faces 14, 15, 23, 28, and 32), and in which classification confidence is at least moderate, yet recognition probability is very low (Faces 13, 17, 19, 22, 29, and 31). The apparent dissociation exhibited in Figure 4 poses an interesting challenge to a model that purports to account quantitatively for

Table 1
Observed and Predicted Category 1 Response Probabilities and "Old" Recognition Probabilities for Each Face in Experiments 1A and 1B

Face	Experiment 1A, all subjects		Experiment 1A, learners only		Experiment 1B, all subjects		Face	Experiment 1A, all subjects		Experiment 1A, learners only		Experiment 1B, all subjects	
	$P(C_1)$	P(old)	$P(C_1)$	P(old)	$P(C_1)$	P(old)		$P(C_1)$	P(old)	$P(C_1)$	P(old)	$P(C_1)$	P(old)
1	.915	.696	.976	.683	.885	.615	18	.148	.793	.082	.780	.110	.752
	.968	.778	.975	.775	.956	.669		.145	.868	.113	.825	.094	.769
2	.716	.833	.821	.825	.619	.764	19	.385	.268	.191	.288	.379	.179
	.665	.741	.850	.700	.594	.756		.281	.281	.138	.250	.325	.138
3	.951	.882	.975	.876	.910	.859	20	.617	.713	.816	.688	.529	.631
	.937	.924	.987	.911	.931	.838		.780	.635	.848	.570	.681	.669
4	.963	.878	.987	.873	.937	.841	21	.241	.391	.112	.401	.123	.312
	.975	.918	1.000	.900	.944	.856		.291	.335	.167	.359	.125	.300
5	.918	.854	.985	.842	.870	.825	22	.318	.340	.162	.362	.298	.257
	.875	.781	.963	.825	.775	.731		.277	.327	.100	.363	.269	.231
6	.106	.722	.034	.705	.071	.674	23	.532	.735	.602	.707	.428	.670
	.134	.675	.075	.663	.100	.588		.601	.741	.538	.700	.450	.600
7	.192	.727	.121	.712	.072	.753	24	.206	.631	.093	.630	.163	.602
	.261	.682	.138	.700	.100	.719		.289	.698	.150	.763	.219	.644
8	.181	.759	.143	.739	.159	.698	25	.344	.559	.214	.546	.333	.506
	.171	.665	.087	.588	.150	.706		.272	.551	.139	.608	.275	.413
9	.089	.726	.082	.719	.066	.666	26	.578	.446	.308	.442	.474	.360
	.089	.753	.050	.700	.069	.738		.490	.503	.238	.600	.413	.431
10	.048	.663	.026	.654	.023	.609	27	.243	.575	.115	.577	.215	.530
	.045	.618	.025	.557	.006	.631		.261	.707	.150	.725	.275	.588
11	.578	.446	.308	.442	.474	.360	28	.328	.789	.383	.762	.210	.791
	.529	.497	.291	.481	.400	.406		.386	.810	.350	.800	.144	.813
12	.631	.383	.731	.393	.443	.316	29	.322	.292	.120	.310	.301	.208
	.538	.247	.718	.256	.488	.263		.269	.300	.175	.275	.294	.169
13	.222	.296	.077	.324	.184	.220	30	.148	.793	.082	.780	.110	.752
	.258	.283	.127	.329	.275	.213		.114	.842	.075	.825	.094	.781
14	.518	.670	.276	.640	.399	.661	31	.895	.415	.957	.437	.839	.316
	.541	.730	.275	.725	.425	.694		.796	.350	.937	.329	.806	.400
15	.385	.743	.175	.724	.229	.750	32	.552	.808	.563	.794	.411	.746
	.430	.753	.200	.713	.238	.719		.538	.848	.544	.810	.300	.719
16	.245	.601	.112	.584	.221	.572	33	.948	.901	.980	.892	.911	.877
	.283	.579	.100	.638	.269	.563		.981	.943	.988	.925	.931	.906
17	.783	.347	.834	.357	.675	.278	34	.133	.814	.062	.799	.085	.795
	.778	.253	.859	.218	.675	.263		.126	.862	.087	.875	.050	.844

Note. Top entries in each row are predicted probabilities; bottom entries are observed probabilities.

both the classification and recognition data within a unified framework.

Exemplar-similarity model. The formal approach to predicting the classification and recognition data is similar to one used in previous work (Nosofsky, 1986, 1987, 1988a), with some technical differences to be explained shortly. The probability that face i is classified in Category 1 is predicted as follows. First, the distance between faces i and j is computed using a (weighted) Euclidean distance metric (e.g., Carroll & Wish, 1974):

$$d_{ij} = c[\sum_m w_m |x_{im} - x_{jm}|^2]^{1/2}, \qquad (1)$$

where x_{im} is the psychological value of face i on dimension m; w_m ($0 \leq w_m \leq 1$, $\sum w_m = 1$) is the weight given to dimension

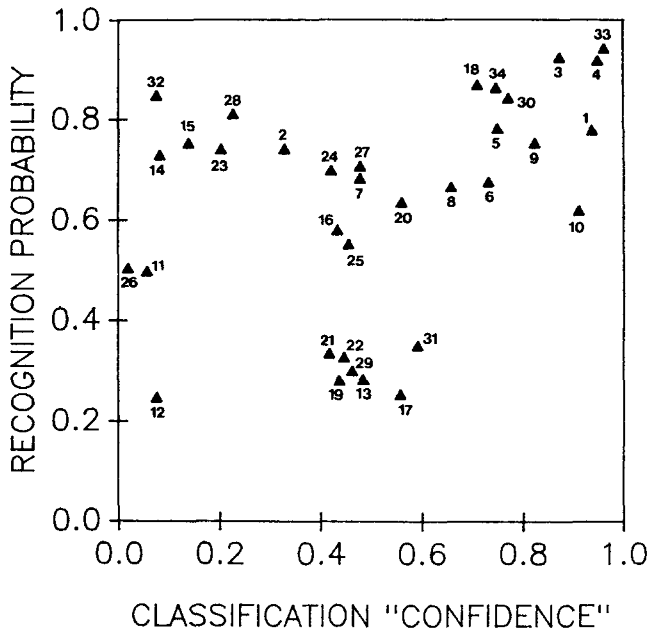

Figure 4. Scatterplot of observed recognition probabilities against observed classification confidences in Experiment 1A.

m; and c is a sensitivity parameter reflecting overall discriminability in the psychological space. The x_{im} psychological values are obtained from the MDS solution derived from the similarity-ratings data (see Appendix A). The weights (w_m) are free parameters to be estimated from the data, and are interpreted as reflecting the attention given to each dimension in making classification judgments.[2]

The distance between faces i and j is transformed to a similarity measure using an exponential decay function (Shepard, 1958, 1987):

$$s_{ij} = \exp(-d_{ij}). \tag{2}$$

The empirical support for an exponential relation between similarity and psychological distance is so pervasive that Shepard (1987) proposed it as a candidate for a universal law of psychological generalization and developed a cognitive process model to account for the law. As explained in previous work, the nonlinear relation between similarity and psychological distance allows the exemplar model to be sensitive to correlational structure in categories, category density effects, and context (Medin & Reynolds, 1985; Medin & Schaffer, 1978; Nosofsky, 1984, 1987, 1988a).

In the present development, it is assumed that the degree to which face i activates exemplar j in memory (a_{ij}) is given by

$$a_{ij} = s_{ij} + e_j, \tag{3}$$

where the e_js are independent and identically distributed normal random variables with mean zero and variance σ^2. The evidence for Category 1 given presentation of face i is found by summing the activations of face i to all exemplars of Category 1,

$$E_{1,i} = \sum_{j \in C1} a_{ij}, \tag{4}$$

and likewise for the evidence for Category 2.

According to the classification model, a Category 1 response is made if the evidence for Category 1 exceeds the evidence for Category 2 by a criterial amount:

$$E_{1,i} - E_{2,i} > b, \tag{5}$$

where b is a response-bias parameter. Note that what is important for classification is the relative activations for the respective categories, that is, the magnitude of $E_{1,i}$ in relation to $E_{2,i}$.

For recognition, the decision rule is to respond "old" if the summed activation for both categories exceeds a criterion x_c:

$$E_{1,i} + E_{2,i} > x_c. \tag{6}$$

Note that what is important for recognition is the overall summed activation for the two categories, not their activations in relation to one another.

As explained in Appendix B, the predicted classification and recognition probabilities are derived from the decision rules in Inequalities 5 and 6 by using a numerical approximation to the integrals of appropriate normal density functions. The free parameters in the model are the overall sensitivity parameter (c) and the attention weights (w_m) in the distance function (Equation 1), the error variance (σ^2) associated with the random variables e_j, the category response-bias parameter (b), and the recognition criterion (x_c). In all of the model-fitting analyses, the sensitivity parameter (c) and error variance (σ^2) were held constant across classification and recognition; however, in the initial analyses the attention weights (w_m) were allowed to vary across classification and recognition, for reasons to be explained shortly. Note that only three of the four attention weights are free parameters, because the weights are constrained to sum to one.

The main (technical) difference between the present formal approach and the previous one adopted by Nosofsky (1986, 1988a) is that in the previous work, the category activation functions were deterministically related to the exemplar similarities, and a probabilistic decision rule was used. By contrast, in the present approach the category activation functions are random variables (Equations 3 and 4), and deterministic decision rules are used (Inequalities 5 and 6). I adopted the latter approach in the present work because I considered it to have a more natural process interpretation. In addition, evidence provided by Ashby and Gott (1988) and Ashby and Maddox (1990) indicates that people often adopt deterministic classification decision rules. Although the present approach involves the use of an additional free parameter (namely, the error variance σ^2), it appears to be necessary if the context model is to adequately describe performance in situations in which people use deterministic decision rules. There are numerous possible sources of the noise in the

---

[2] In deriving the initial MDS solution from the similarity-ratings data, the weights in the distance function are nonidentifiable with respect to the MDS coordinates. It is assumed implicitly that in making their similarity judgments subjects gave roughly equal attention weight to each of the dimensions, which seems reasonable in a context-free similarity-judgment task.

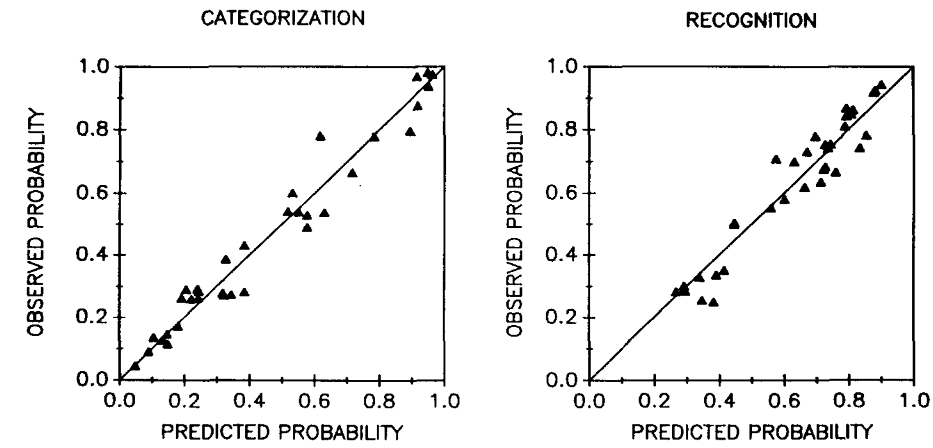

Figure 5. Scatterplot of observed against predicted classification and recognition probabilities in Experiment 1A.

classification and recognition process, including noise in the locations of the exemplars in the multidimensional space (e.g., Ennis, Palen, & Mullen, 1988; Zinnes & MacKay, 1983), noise in the similarity computations, and noise in the criterion settings (e.g., Wickelgren, 1968), but the issue of the precise locus of the noise is not pursued in this research.

Quantitative fits. The exemplar model (Equations 1–6) was fitted simultaneously to the classification and recognition data with maximum likelihood as the criterion of fit.[3] The observed classification and recognition probabilities are plotted against the predicted probabilities in Figure 5. (The predicted probabilities are also reported in Table 1.) It is evident from inspection that the model provides good fits to both sets of data, accounting for 95.9% of the variance in the classification probabilities and for 91.9% of the variance in the recognition probabilities.

This analysis illustrates the benefits that can accrue from a formal mathematical modeling approach. The initial scatterplot in Figure 4 revealed little relation between classification and recognition performance. At that limited level of analysis, one might have concluded that there was little in common between the fundamental processes of classification and recognition. Under the guidance of the formal model, however, a unified account of these processes is achieved in terms of the assumption that classification and recognition are governed by similarity comparisons to stored exemplars.

To gain insight into the workings of the model, consider again the MDS solution for the exemplars in Figure 3 and the classification–recognition scatterplot in Figure 4. Faces 13, 19, 22, and 29 are examples of faces that received low recognition probabilities, yet for which there was moderate classification confidence. Inspection of the MDS solution reveals that these faces are located in an isolated region of the psychological space (the upper-right region of the plot of Dimension 1 vs. Dimension 2). Thus, the summed similarity for these faces is low, so the model correctly predicts their low recognition probabilities. In a relative sense, however, these faces are clearly more similar to the exemplars of Category 2 than to the exemplars of Category 1, so the model correctly predicts that these faces will be classified into Category 2 with moderately high probability. Conversely, Faces 14, 15, 23, 28, and 32 received high recognition probabilities but had low classification confidence. These faces tended to be centrally located in the MDS solution and so were moderately similar to numerous exemplars. In addition, they were highly proximal to one of the training exemplars (either Face 2 or Face 7). Thus, the overall summed similarity for these faces was high, explaining their high recognition probabilities. Both categories of exemplars exerted competing influence for these faces, however, so there was low classification confidence associated with them. Finally, there were also items such as Faces 3, 4, and 33, which were highly similar to several exemplars in the MDS solution from only the same category. For these faces, both absolute summed similarity and relative target-to-contrast similarity were high, explaining jointly their high recognition probabilities and classification confidences. The examples just discussed provide formal illustrations of conjectures concerning the classification–recognition relation put forth previously by Medin (1986).

In subsequent sections of this article, it will be useful to consider performance patterns exhibited by learners separately from the grouped-subjects data. The *learners* are de-

[3] A computer search was used to find the parameters that maximized the following log-likelihood function: $\ln L = \sum_{i=1}^{34} \ln N_i! - \sum_{i=1}^{34} \sum_{j=1}^{2} \ln f_{ijC}! + \sum_{i=1}^{34} \sum_{j=1}^{2} f_{ijC} \ln P_{ijC} + \sum_{i=1}^{34} \ln N_i! - \sum_{i=1}^{34} \sum_{j=1}^{2} \ln f_{ijR}! + \sum_{i=1}^{34} \sum_{j=1}^{2} f_{ijR} \ln P_{ijR}$, where N_i is the frequency with which stimulus i was presented; f_{ijC} is the observed frequency with which stimulus i was classified in category j; f_{ijR} is the observed frequency with which stimulus i was judged in the recognition task as either "old" ($j = 1$) or "new" ($j = 2$); and p_{ijC} and p_{ijR} are the corresponding predicted classification and recognition probabilities, respectively. This likelihood function assumes that the classification and recognition probabilities for each stimulus are binomially distributed. Because the likelihood function also assumes that the distributions for each stimulus are independent (an assumption that is probably wrong given that multiple observations are collected from the same subjects), the results of the statistical tests should be interpreted with caution.

fined as those subjects whose performance during the second half of the training phase fell above the 50th percentile point, where performance is defined in terms of average percentage correct. (In the present experiment, the learners averaged 95.1% correct during the second half of training, whereas the nonlearners averaged 81.1% correct.) Table 1 reports the classification and recognition data obtained during the transfer phase for the learners only. As was the case for the full set of subjects, there was a low correlation between recognition probability and classification confidence (r = .12), yet the exemplar model accounted for an impressive 98.9% of the variance in the classification data and for 85.1% of the variance in the recognition data. The predicted classification and recognition probabilities for the learners are presented alongside the observed data in Table 1.

Selective attention in classification and recognition. A cornerstone of the context model is the assumption that similarities between exemplars are modifiable by selective attention processes (Medin & Schaffer, 1978; Nosofsky, 1986). Selective attention is modeled by the dimension-weight parameters in the distance function (Equation 1). Previous research has demonstrated systematic shifts in the magnitudes of the attention weights depending on the structure of the categories that are learned. In particular, there is support for the hypothesis that subjects may distribute their attentional resources over the component dimensions of the stimuli to optimize their classification performance (Getty, Swets, Swets, & Green, 1979; Nosofsky, 1984, 1986, 1987, 1989; Reed, 1972; Shepard et al., 1961).

A question of interest in the present research is whether the attention weights may vary depending on whether subjects are making classification or recognition judgments. It is critical to realize that dimensions that are highly diagnostic for making classification judgments may be useless for distinguishing between old and new items, and vice versa. To take an extreme example, suppose that all exemplars presented during the training phase were colored blue, and all new items presented during transfer were colored red. The color of an item would be useless for purposes of classifying it into Category 1 or 2, and subjects would presumably give it zero weight. However, subjects could use the color dimension to discriminate perfectly between the old and new items and so might focus most of their attention on color for purposes of recognition.

In the present experiment, subjects were trained to classify the stimuli and received explicit classification feedback. Thus, the hypothesis that subjects may have fine-tuned their attentional weightings during training to optimize classification performance is plausible. By contrast, there was obviously no recognition training, and no explicit recognition feedback was provided. Furthermore, the strategies of attention weighting adopted during classification training may constrain what subjects are able to do with regard to recognition. For example, if a dimension is not attended during classification training, impoverished information about the values on that dimension may be stored in memory. And if the information is not stored adequately in memory, subjects clearly would not be able to take full advantage of that dimension when asked later to make recognition judgments. For the aforementioned reasons, it seems implausible that subjects could distribute attention over the psychological dimensions to optimize recognition performance. Nevertheless, it seems reasonable to explore the weaker hypothesis that subjects may modify their attentional weightings to improve recognition performance in relation to what would be accomplished if they used the same weights for recognition as are used for classification.

The maximum likelihood parameters and summary fits for the context model are reported in Table 2, separately for the all-subjects analyses and the learners-only analyses. Inspection of the parameters suggests that the distributions of attention over the psychological dimensions were indeed different for classification and recognition. To corroborate this observation, a restricted version of the context model was fitted to the data in which the attention weights were constrained to be constant across classification and recognition.[4] The fits for this restricted model, summarized in Table 2, were noticeably worse than for the full model. For example, for the learners the sum of squared deviations between predicted and observed classification and recognition probabilities increased from .048 and .222, respectively, to .129 and .367. Likelihood ratio tests indicated that the restricted model fit significantly worse than the full model, $\chi^2(3, N = 10{,}880) = 125.6$, $p < .001$, for the all-subjects analysis, and $\chi^2(3, N = 10{,}880) = 95.4$, $p < .001$, for the learners-only analysis, so there is statistical support for the hypothesis of varying attention weights.

To investigate whether the different distributions of attention could be interpreted in terms of the attention-optimization hypothesis, a computer search was conducted to find the optimal weights for classification and recognition. For classification, the optimization criterion was defined as the maximum average percentage of correct classifications that could be achieved, whereas for recognition it was defined as the maximum average hit rate minus false-alarm rate. In conducting these searches, all parameters with the exception of the attention weights were held fixed at those values that yielded a maximum likelihood fit to the empirical classification and recognition data.

The performance-optimizing (or *ideal-observer*) distributions of attention weights are plotted along with the best fitting distributions in Figure 6 for the learners only. With regard to classification, the correspondence between the ideal-observer and the best fitting distributions of weights is impressive, with primary weight given to nose length, secondary weight given to eye height, and virtually no weight given to eye separation or mouth height. (Intuitively, the reason it is optimal to attend to eye height and nose length in the present experiment is because the combination of these dimensions yields the largest separation between the two categories of

[4] A restricted version of a model arises when some of its parameters are constrained on a priori grounds. Let $\ln L(F)$ and $\ln L(R)$ denote the log-likelihoods for a full and restricted model, respectively. Assuming the restricted model is correct, the quantity $-2*[\ln L(R) - \ln L(F)]$ is distributed as a chi-square random variable with degrees of freedom equal to the number of constrained parameters. If this quantity exceeds the critical value of chi-square, then one would conclude that some of the parameters were constrained inappropriately. (See Wickens, 1982, Chapter 5, for a complete exposition of likelihood-ratio testing.)

Table 2
Maximum Likelihood Parameters and Summary Fits, Experiment 1A

	Parameters								Fits		
Model	σ	c	w_1	w_2	w_3	w_4	x_c	b	SSE	% Var	-ln L
					All-subjects analyses						
Context											
Classification	.289[a]	1.177[a]	.19	.12	.25	.45		.181	.115	95.9	136.0
Recognition	.289[a]	1.177[a]	.17	.48	.25	.10	4.498		.128	91.9	141.0
Restricted context											
Classification	.337[a]	1.617[a]	.24[a]	.29[a]	.30[a]	.18[a]		.108	.176	93.7	161.8
Recognition	.337[a]	1.617[a]	.24[a]	.29[a]	.30[a]	.18[a]	3.198		.227	85.5	178.0
Prototype											
Classification	.230[a]	.902[a]	.22	.14	.40	.24		−.008	.215	92.3	191.8
Recognition	.230[a]	.902[a]	.24	.50	.18	.08	1.105		.205	86.9	168.0
					Learners-only analyses						
Context											
Classification	.319[a]	1.360[a]	.29	.00	.64	.07		.162	.048	98.9	82.3
Recognition	.319[a]	1.360[a]	.15	.46	.26	.13	4.006		.222	85.2	118.6
Restricted context											
Classification	.308[a]	1.661[a]	.32[a]	.14[a]	.47[a]	.06[a]		.169	.129	97.0	102.8
Recognition	.308[a]	1.661[a]	.32[a]	.14[a]	.47[a]	.06[a]	3.203		.367	75.5	145.9
Prototype											
Classification	.214[a]	.923[a]	.30	.03	.61	.07		.005	.205	95.2	133.6
Recognition	.214[a]	.923[a]	.27	.43	.21	.09	1.089		.212	85.9	116.2

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood. Criterion of fit was to minimize -ln L; SSE and % Var are auxiliary measures.
[a] These parameters were held fixed across classification and recognition.

exemplars in the multidimensional space. The interested reader can verify this point by seeing Figure 9B and noting the large separation between Faces 1–5 and Faces 6–10 in the plot of eye height versus nose length.)

As might be expected given the absence of explicit training, the correspondence between the ideal-observer and the best fitting distributions of attention weights is not very good for recognition. (Alternative optimization criteria besides maximizing average hit minus false-alarm rate were also considered, but none yielded a close correspondence between the ideal-observer and the best fitting weights.) It is of interest to note, however, that whereas subjects gave little weight to eye separation in making their classification decisions, a good deal of weight was shifted to this dimension for recognition, which is precisely as would be predicted by the attention-optimization hypothesis. Indeed, according to the model subjects would have achieved a hit minus false-alarm rate of only .087 had they used the same weights in the recognition task as they had used in the classification task. By way of comparison, the hit minus false-alarm rate predicted by the model with the best fitting recognition weights is .192. Thus, subjects improved their recognition performance substantially by shifting some attention to the eye-separation dimension. (With the optimal recognition weights, subjects could have achieved a hit minus false-alarm rate of .338.)

In summary, there is support for the hypothesis that the learners fine-tuned their attention weights to optimize classification performance, and there were tendencies in that direction with regard to recognition. Inspection of the best fitting attention weights for the all-subjects data reveals that they are far from the ideal-observer distribution. One interpretation is that part of the nonlearners' difficulty in solving the classification problem is that they focused attention on nondiagnostic dimensions.

Comparisons with a prototype model. The purpose of this experiment was to investigate the classification–recognition relation within the framework of the exemplar-based context model. The category structure that was tested was not designed to yield qualitative contrasts between the predictions of the context model and those of prototype models. Nevertheless, it is of interest to compare the quantitative predictions of the competing models, particularly because Reed (1988) concluded that the prototype model is superior in situations in which people classify continuous-dimension stimuli such as dot patterns and schematic faces (but see Ashby & Gott, 1988; Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Nosofsky, 1986, 1987, 1988a, 1988c).

First, note that the physical prototypes (Faces 33 and 34) were classified with extremely high accuracy during transfer. Although consistent with a prototype model, this result is also accurately predicted by the context model (see Table 1). The context model makes this prediction because the prototypes are highly similar to the exemplars of their own category and are highly dissimilar to the exemplars of the contrast category.

Other transfer stimuli of interest are what Reed (1972) termed the *control faces* (Faces 31 and 32). According to Reed, these faces were equated with the prototypes in terms of their similarity to the exemplars of the alternative categories. Because the prototypes were classified with higher accuracy than the control faces, Reed argued that the results favored the prototype model over exemplar-based models. This conclusion is problematic, however, because as is seen

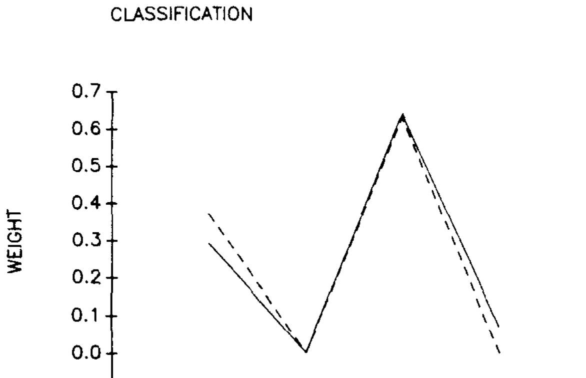

Figure 6. Best fitting distributions of attention weights (solid lines) plotted with ideal-observer distributions of attention weights (dashed lines), Experiment 1A.

in Table 1, the exemplar-based context model accurately predicts classification probabilities for the control faces. Indeed, the control face that purportedly belongs to Category 2 (Face 32) is accurately predicted by the context model to be classified with slight tendency in Category 1. Although the control faces may be equated with the prototypes on some type of similarity measure, they are clearly not equated with respect to the predictions of the context model. To summarize, the context model accurately predicts classification performance for the prototypes and control faces used by Reed (1972).

To make quantitative comparisons, prediction equations analogous to Equations 1–6 were formulated for a prototype model (see Appendix C for details). The free parameters for the prototype model and the context model were identical and functioned analogously. The only difference between the models is that instead of computing summed similarity between a probe and the category exemplars, the prototype model computes the similarity between a probe and the category central tendency (i.e., the psychological prototypes illustrated in Figure 3). The maximum likelihood parameters and summary fits for the prototype model are reported in Table 2. For the all-subjects analyses, the quantitative predictions of the prototype model are clearly worse than those achieved by the exemplar-based context model for both classification and recognition. For the learners-only analyses, the context model outperforms the prototype model on the classification data, and the fits are essentially the same for the recognition data. These results lead one to strongly question Reed's (1988, p. 172) assertion that prototype models are superior to exemplar models in predicting classification in continuous-dimension stimulus domains.

Role of the configural property. In light of the finding that the learners gave primary weight to the dimensions of eye height and nose length in making their classification judgments, a question that arises is the extent to which they were relying on the configural property of distance between the level of the eyes and the top of the nose. Recall that this configural property is a perfect predictor of category membership (see Figure 1). Similarity judgments based solely on eye–nose distance look very different from ones based on eye height and nose length treated as separate dimensions. For example, if eye–nose distance were the only relevant dimension, then a face with high eyes and a long nose would be highly similar to one with low eyes and a short nose. By contrast, if eye height and nose length operated as individual psychological dimensions, these same two faces would be highly dissimilar (as was observed in the similarity-ratings data). Thus, there is a clear distinction between an exemplar-similarity model which assumes that eye–nose distance is the sole underlying psychological dimension and one that treats eye height and nose length as individual dimensions.

Numerous theoretical analyses of the learners' classification and recognition data were conducted to test for the influence of the configural property. In one set of analyses, a fifth psychological dimension was defined by setting $x_{i5} = a * x_{i1} - b * x_{i3}$, where x_{i5} is an item's coordinate on the dimension of eye–nose distance and a and b are scaling parameters. In another set of analyses, the physical distance between level of eyes and top of nose was measured for each face, and the psychological distance was assumed to be linearly related to this physical distance. Regardless of its definition, use of this fifth dimension in addition to the original four dimensions led to no improvement in the quantitative fit that could be achieved, and the weight given by the exemplar model to the fifth dimension was essentially zero. In other analyses, the fifth dimension was used instead of Dimensions 1 and 3 (i.e., the weights for Dimensions 1 and 3 were held fixed at zero). In this case the exemplar model yielded a quantitative fit to the learners' classification and recognition data that was worse than the one achieved previously (total $-\ln L = 241.2$; classification sum of squared deviations [SSE] = .096, recognition SSE = .349). Although these analyses do not rule out the hypothesis that some subjects may have made extensive use of the configural property, it does not appear to provide the sole or primary account of the present data. Regardless of the extent to which subjects treated eye height and nose length as individual dimensions as opposed to treating them in a configural manner, the exemplar model provides an excellent characterization of the classification and recognition data.

Summary

This experiment provided support for the exemplar-based approach to predicting and relating perceptual classification and recognition. According to the model, classification and recognition decisions are based on similarity comparisons to stored exemplars; but whereas classification judgments involve a relative-similarity rule, recognition judgments involve an absolute summed-similarity rule. In addition to accounting for a low correlation between recognition probabilities and classification confidence, the model proved capable of predicting fine-grained differences in these judgments on the basis of fine-grained differences in similarities among items. Evidence was also provided that the learners fine-tuned their attentional weightings of the component dimensions of the objects to optimize their classification judgments, and some tendencies in that direction were observed for recognition. Finally, the exemplar model accurately predicted prototype phenomena in the classification and recognition data and was superior to a central-tendency prototype model in quantitative accuracy.

Experiment 1B

One of the fundamental variables affecting learning and memory is frequency or repetition. The purpose of Experiment 1B was to explore the role of individual item frequency in perceptual classification and recognition and to use the frequency manipulations to provide further tests of the exemplar-based approach to modeling performance in these tasks. Estes (1986a, 1986b) and Nosofsky (1988c) tested a version of the context theory which assumed that repeating an exemplar during classification learning leads to strengthening the representation of that exemplar in memory. In these previous theoretical investigations, the exemplar memory strengths were assumed to be directly proportional to the relative frequencies with which the exemplars were presented during training, as in a pure multiple-trace model of memory (e.g., Hintzman, 1986, 1988; Hintzman & Block, 1971). I weaken this assumption in the present work by allowing the memory strength for a high-frequency item to be a free parameter. In the present development, the degree to which item i activates exemplar j in memory is given by

$$a_{ij} = M_j s_{ij} + e_j, \tag{7}$$

where M_j is the memory strength associated with exemplar j, s_{ij} is the similarity between items i and j, and e_j is the random variable defined earlier. Thus, the degree to which exemplar j is activated is a joint function of its strength in memory, its similarity to the presented item i, and random noise.

In Experiment 1B, Face 7 from Category 2 was presented five times as often as any of the other exemplars during the training phase (see Figure 3). I decided to vary the presentation frequency of an atypical face to avoid potential ceiling effects on the frequency manipulation. The critical predictions stemming from the frequency-sensitive exemplar model are that in relation to the Experiment 1A baseline condition, Category 2 response probabilities and "old" recognition probabilities associated with Face 7, and stimuli that are highly similar to Face 7, will increase. Little effect of the frequency manipulation is predicted for items that are dissimilar to Face 7. That is, the model (Equation 7) predicts an interactive effect of similarity and frequency. A good metaphor is that Face 7 will act as a magnet in the psychological space, drawing nearby stimuli toward it.

In experiments involving the classification of Munsell colors, Nosofsky (1988c) reported previous tests of the frequency-sensitive exemplar model that supported these qualitative predictions. The present experiment extends this earlier work in several respects. First, Nosofsky's (1988c) demonstrations were limited to showing that manipulations of exemplar frequencies and similarities significantly influenced typicality and confidence judgments, whereas the focus in the present experiment is on actual choice probabilities. Second, whereas the previous design included only training items, the present experiment tests how frequency and similarity interact to determine generalization to transfer stimuli. Finally, by also testing in the present experiment how frequency and similarity influence people's recognition judgments, further constraints are placed on the proposed exemplar approach to relating classification and recognition.

Method

Subjects

The subjects were 80 undergraduates from Indiana University who either participated for course credit or were paid. All subjects were tested individually.

Stimuli and Apparatus

The stimuli and apparatus were the same as in Experiment 1A.

Procedure

The procedure was the same as in Experiment 1A, except that Face 7 was presented five times as often as any of the other faces during the training phase. There were 12 blocks of training trials, with 14 trials per block. Face 7 was presented five times in each block, and the remaining nine faces were presented once each. Assignment of faces to trial numbers was randomized within each block.

Results

Effects of Exemplar Frequency

To facilitate discussion, Experiment 1A will be referred to as the equal-frequency condition (EF) and Experiment 1B as the high-frequency-7 condition (HF7). The probability with which each face was classified in Category 1 and was recognized as "old" during the transfer phase of Condition HF7 is reported in Table 1. For convenience in making comparisons. Figure 7 plots the classification and recognition probabilities observed in Condition HF7 against those observed in Condition EF. In the figure, the solid circles are used to highlight critical stimuli for which the frequency manipulation was expected to have large effects, namely, Faces 7, 14, 15, 21, and 28. Face 7 was the frequency-manipulated stimulus, whereas Faces 14, 15, 21, and 28 were highly similar to Face 7, differing from it on only one (physical) dimension and matching it on the remaining three dimensions (see Figure 3). If there were no changes in classification or recognition probabilities across Conditions EF and HF7, then all points plotted in Figure 7 would lie on the diagonal. The effect of the frequency manipulation can be viewed by noting the direction and extent to which the points depart from the diagonal.

Regarding the classification results, the Category 1 response probabilities show clear and sizable decrements in Condition HF7 in relation to Condition EF for the critical stimuli highlighted in Figure 7A. Furthermore, as predicted by the exemplar model, the frequency manipulation interacted with similarity in influencing the classification probabilities. It was not simply the case that all classification probabilities in Condition HF7 were lowered substantially in relation to the ones observed in Condition EF; rather, the effect of the frequency manipulation depended on stimulus. To demonstrate this point, a difference score was calculated for each stimulus by subtracting the Category 1 response probability in Condition EF from the one in Condition HF7. For the five critical stimuli the average difference score was $-.174$, whereas for the remaining 29 noncritical stimuli the average difference score was only $-.046$, $t(32) = -4.42$, $p < .001$. Nevertheless, Figure 7A shows clear effects of the frequency manipulation on stimuli other than the five critical faces. As will be seen in ensuing theoretical analyses, the exemplar model does indeed predict many of these effects. Most notably, large effects were also observed for Faces 11, 26, and 32 (average difference score $= -.150$), which correspond to the cross-hatched circles in Figure 7A. With the exception of the five critical faces, Faces 11, 26, and 32 were closest in (weighted) distance to Face 7 in the psychological space. (The only sizable misprediction across the 34 faces was for Face 23, which had a larger than predicted difference score.)

Regarding the recognition results, Figure 7B shows a general lowering of the HF7 probabilities in relation to the EF probabilities, an effect that is interpretable in terms of an overall criterion shift. The important question is whether frequency interacted with stimulus similarity in influencing the recognition judgments. A difference score was calculated for each stimulus by subtracting the recognition probability in Condition EF from the one in Condition HF7. For the five critical stimuli the average difference score was $-.014$, whereas for the remaining 29 stimuli the average difference score was $-.057$. This result is in the predicted direction but falls short of statistical significance, $t(32) = -1.53$, $.05 < p < .10$ (one-tailed test).

A potential problem with the preceding analyses is that changes in probability are not equivalent at all regions of the scale. Because the probability measure has a floor and ceiling (at 0.0 and 1.0), a change in probability, say, from .10 to .05 is clearly more significant than a change from .50 to .45. To remove this potential confounding, new analyses were conducted in which the individual probabilities were transformed to z scores. (The z transformation has the effect of stretching the probability scale at its edges.) Then, as in the previous analyses, differences in z scores between Conditions EF and HF7 were computed. For classification, the average z-score difference for the five critical stimuli was $-.564$ and for the 29 remaining stimuli it was $-.192$, $t(32) = -3.69$, $p < .01$. For recognition, the average z-score difference for the five critical stimuli was $-.040$ and for the remaining 29 stimuli it was $-.191$, $t(32) = 1.72$, $p < .05$ (one-tailed test).

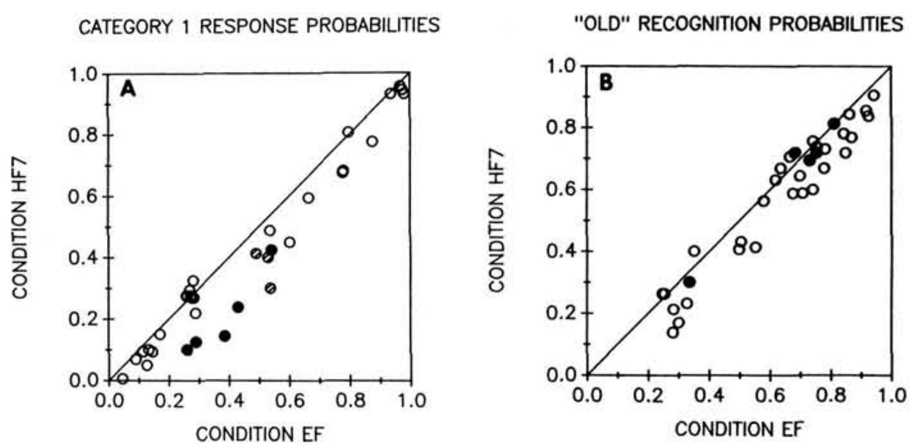

Figure 7. Observed Category 1 response probabilities for Condition HF7 plotted against observed Category 1 response probabilities for Condition EF (Panel A), and observed "old" recognition probabilities for Condition HF7 plotted against observed "old" recognition probabilities for Condition EF (Panel B). (In both panels, the results for the 5 critical stimuli [the high-frequency item and its neighbors] are marked by the solid circles, whereas the results for the 29 noncritical stimuli are marked by the open circles. The cross-hatched circles mark results for stimuli that are proximal to the high-frequency item but not as proximal as the critical stimuli.)

In summary, the difference-score analyses using both the raw probability data and the z-transformed data support the exemplar model's predictions of an interactive effect of frequency and similarity, although the effect on recognition is clearly weaker than the effect on classification. Discussion of possible reasons for the relatively weak effect on recognition is withheld until after the collection of additional data in Experiment 2.

Theoretical Analyses

The frequency-sensitive exemplar model (Equations 1, 2, 4–7) was fitted to the classification and recognition data obtained in Condition HF7 with maximum likelihood as the criterion of fit. The memory strength for Face 7 (M_7 in Equation 7) was a free parameter, whereas the memory strengths for the remaining nine training faces were set at 1.0. (The strength parameter for Face 7 was held fixed across classification and recognition.) The observed classification and recognition probabilities are plotted against the predicted probabilities in Figure 8, and the predicted probabilities are also reported in Table 1. As was the case in Experiment 1A, the exemplar model provides good fits to both sets of data, accounting for 96.5% of the variance in the classification probabilities and for 95.4% of the variance in the recognition probabilities. Once again, these accurate quantitative fits were achieved despite the fact that the correlation between recognition and classification confidence was low, $r = .50$. The model was also fitted to the learners-only data, that is, the data for those subjects who were the top 40 performers during the second half of the training phase. For these subjects, the exemplar model accounted for 98.3% of the classification-probabilities variance and for 93.3% of the recognition-probabilities variance.

The maximum likelihood parameters and summary fits for the exemplar model are reported in Table 3. The patterns of attention weights closely parallel the ones observed in Experiment 1A. Again, for the learners, the best fitting distribution of attention weights comes close to the ideal-observer distribution, and there were tendencies in that direction for recognition.

A frequency-sensitive prototype model was also tested. The prediction equations were the same as in Appendix C, except that in calculating the central tendency for Category 2, a weighted average over the exemplars was used. Specifically, the weight given to Face 7 in calculating the central tendency was allowed to be a free parameter. To the extent that Face 7 is highly weighted, the central tendency shifts in its direction. The maximum likelihood parameters and summary fits for the frequency-sensitive prototype model are reported in Table 3. Again, this model performed worse than did the exemplar model at predicting the classification and recognition probabilities.

Exemplar strength. The central question of interest in this experiment concerned the role of frequency in classification and recognition, and the extent to which the frequency effects could be modeled in terms of changing exemplar strength. As reported in Table 3, the maximum likelihood strength parameter for Face 7 was $M_7 = 1.464$. Use of this free parameter was critical for accurately characterizing the classification and recognition data. A restricted version of the model in which the strength parameter was assumed to be directly proportional to the relative presentation frequency of Face 7 (i.e., $M_7 = 5.0$) performed dramatically worse than the full model, $\chi^2(1, N = 10{,}880) = 516.7$, $p < .001$. The restricted model predicted changes in classification and recognition probabilities for Face 7 and its neighbors that were far too extreme. Thus, there appears to be a negatively accelerated, increasing relation between represented and actual frequency. In other words, once exemplar strength reaches certain levels, additional presentations of an item may lead to diminishing increases in exemplar strength. In part, this negatively accelerated relation may be reflecting a phenomenon in which subjects devote less attention and rehearsal to the highly frequent items that are already well stored in memory.

A frequency-insensitive model with $M_7 = 1.0$ also performed significantly worse than the full model, $\chi^2(1, N = 10{,}880) = 15.1$, $p < .01$. This model failed to account for the

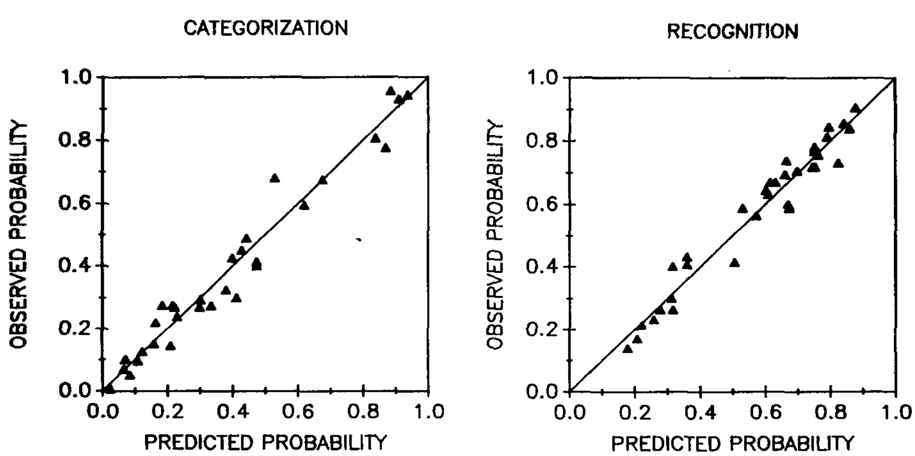

Figure 8. Scatterplot of observed against predicted classification and recognition probabilities in Experiment 1B (Condition HF7).

Table 3
Maximum Likelihood Parameters and Summary Fits, Experiment 1B

Model	Parameters									Fits		
	σ	c	w_1	w_2	w_3	w_4	x_c	b	M_7	SSE	% Var	-ln L
					All-subjects analyses							
Context												
Classification	.267[a]	1.077[a]	.15	.15	.29	.41		.173	1.464[a]	.097	96.5	129.2
Recognition	.267[a]	1.077[a]	.13	.56	.23	.08	5.322		1.464[a]	.076	95.4	119.2
Prototype												
Classification	.186[a]	.777[a]	.16	.14	.40	.30		.044	1.123[a]	.175	93.7	181.0
Recognition	.186[a]	.777[a]	.25	.55	.12	.07	1.231		1.123[a]	.182	89.0	156.0
					Learners-only analyses							
Context												
Classification	.298[a]	1.491[a]	.25	.02	.62	.11		.269	1.544[a]	.073	98.3	81.8
Recognition	.298[a]	1.491[a]	.06	.57	.28	.08	4.430		1.544[a]	.110	93.3	100.1
Prototype												
Classification	.191[a]	.887[a]	.26	.01	.63	.10	1.191		1.386[a]	.114	97.4	112.4
Recognition	.191[a]	.887[a]	.18	.53	.20	.10		.059	1.386[a]	.216	86.8	117.3

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood.
[a] These parameters were held fixed across classification and recognition.

clear interactive effects of frequency and similarity that were observed in the data. When separate exemplar-strength parameters were allowed, the maximum likelihood estimates were $M_7 = 1.583$ for classification and $M_7 = 1.369$ for recognition; however, this model with separate strength parameters did not fit significantly better than the one that assumed constant exemplar strength, $\chi^2(1, N = 10{,}880) = 0.76$, $p > .20$. Thus, although the qualitative effects of the frequency manipulation were weaker for recognition than for classification, the quantitative modeling suggests caution in overinterpreting the result.

To gain additional perspective on the frequency manipulation, theoretical analyses were conducted that directly compared the Conditions EF and HF7 classification and recognition data. First, the exemplar model was fitted simultaneously to the EF and HF7 classification data, with all parameters held constant across conditions except for the exemplar-strength parameter and the category response-bias parameter (b in Inequality 5). As reported in Table 4, impressive fits to both data sets were achieved, with the maximum likelihood estimate of M_7 being 2.028. Thus, the qualitative changes in classification probabilities that resulted from the frequency manipulation are parsimoniously described by an exemplar model with constant sensitivity, attention weight, and variance parameters, with only the exemplar-strength and response-bias parameters varying. When the model was refitted to both sets of classification data with the strength parameter held fixed at $M_7 = 1.0$, the quantitative fit was dramatically worse, $\chi^2(1, N = 10{,}880) = 31.59$, $p < .001$. This model-fitting analysis corroborates the previous qualitative observations made with regard to Figure 7.

The exemplar model was also fitted simultaneously to the EF and HF7 recognition data, with all parameters held constant across conditions except for the exemplar-strength parameter and the recognition criterion (x_c in Inequality 6). Again, impressive fits to both data sets were achieved (Table 4), with the maximum likelihood estimate of M_7 being 1.309. Holding fixed M_7 at 1.0 led to a significantly worse fit, $\chi^2(1, N = 10{,}880) = 6.19$, $p < .05$. However, in agreement with the previous qualitative analyses, this quantitative modeling sug-

Table 4
Maximum Likelihood Parameters and Summary Fits for the Context Model When the Conditions EF and HF7 Classification and Recognition Data Are Fitted Conjointly

Condition	Parameters								Fits		
	σ	c	W_1	W_2	W_3	W_4	b	M_7	SSE	% Var	-ln L
Classification											
EF	.347[a]	1.370[a]	.22[a]	.15[a]	.28[a]	.35[a]	.169	1.000	.108	96.1	133.2
HF7	.347[a]	1.370[a]	.22[a]	.15[a]	.28[a]	.35[a]	−.008	2.028	.094	96.6	132.5
Recognition											
EF	.258[a]	1.117[a]	.15[a]	.48[a]	.25[a]	.11[a]	4.727	1.000	.122	92.2	138.2
HF7	.258[a]	1.117[a]	.15[a]	.48[a]	.25[a]	.11[a]	5.020	1.309	.079	95.2	120.8

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood. The value of M_7 was held fixed at 1.000 in Condition EF.
[a] These parameters were held fixed across Conditions EF and HF7.

gests a weaker interactive effect of frequency and similarity on recognition than on classification.

Discussion

Broadly, the results of Experiment 1B provide additional support for the exemplar approach to relating classification and recognition. Good quantitative fits to both sets of data were achieved, and the fits were superior to those of a central-tendency prototype model. The interactive effect of frequency and similarity on classification and recognition performance was also well captured by the model and was interpretable in terms of changes in exemplar strength resulting from the frequency manipulation. A cause for concern, however, is that the effect of the frequency manipulation was weaker for recognition than for classification, as indicated by both the qualitative and quantitative analyses. Although suggestive, more evidence is needed before considering possible reasons for this unexpected result. The purpose of Experiment 2 was to follow up on the studies of Experiments 1A and 1B by using a new category structure and to obtain additional evidence bearing on the main issues that have been studied.

Experiment 2

A new category structure was designed for testing the exemplar model of classification and recognition. The same set of schematic faces from Experiment 1 was used. The category structure is illustrated in Figure 9, which shows plots of Dimensions 2 versus 4 and Dimensions 1 versus 3. The training exemplars of Category 1 were Faces 12, 19, 22, 27, and 28 (enclosed by circles), whereas the Category 2 training exemplars were Faces 8, 17, 23, 25, and 32 (enclosed by triangles).

As shown in Figure 9A, the categories of exemplars are well separated in the plot of Dimension 2 (eye separation) versus Dimension 4 (mouth height). Indeed, over a wide range of parameter values, the exemplar model predicts that subjects would optimize their classification performance by attending selectively to these psychological dimensions. Thus, whereas for the category structure that was tested in Experiment 1 the learners focused attention on the dimensions of eye height and nose length, it was expected in the present experiment that they would attend primarily to eye separation and mouth height.

Although eye separation and mouth height are highly diagnostic for classification, it turns out that the other two dimensions are more diagnostic for discriminating old items from new items. This point is illustrated in Figure 9B, which shows that the old training exemplars form a diagonal running from the lower left to the upper right of the eye height–nose length psychological space, with many of the new items lying at the opposite corners of this space. Thus, the combination of values on the eye-height and nose-length dimensions allows for a fairly good separation between the sets of old and new exemplars. Indeed, over a wide range of parameter values, the exemplar model predicts that subjects would optimize their recognition performance by attending selectively to eye height and nose length. The central prediction in this experiment, therefore, is that whereas subjects will focus on eye separation and mouth height for classification purposes, there will be some shifts of attention to eye height and nose length for recognition purposes.

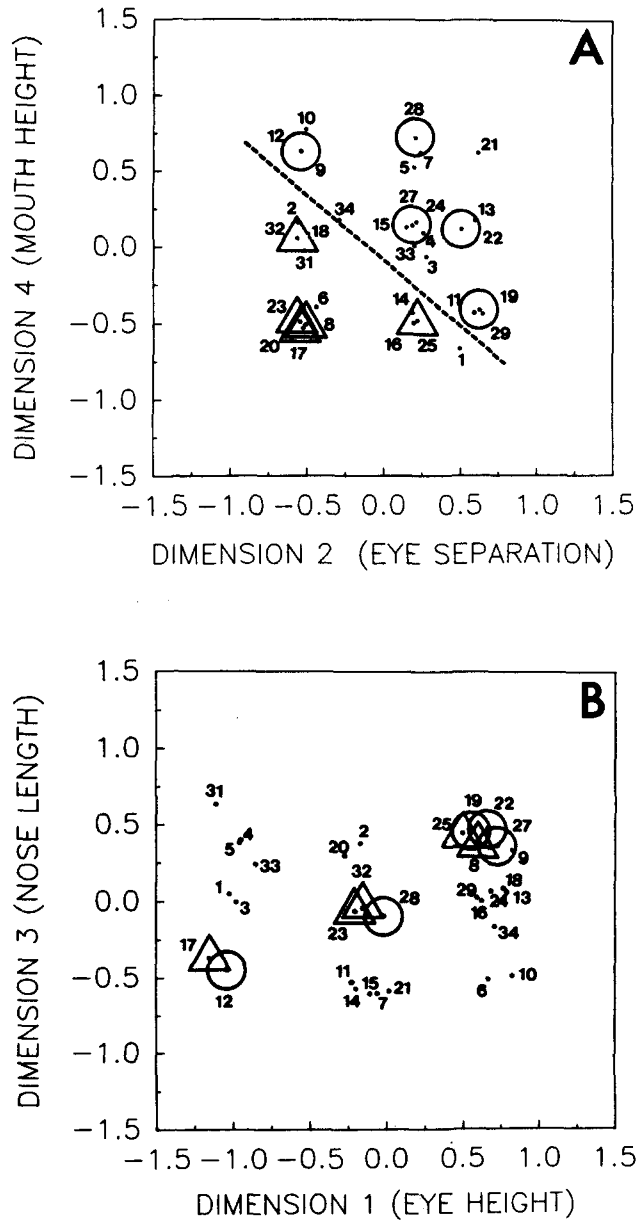

Figure 9. Illustration of the category structure tested in Experiment 1. (Stimuli enclosed by circles denote training exemplars for Category 1, and stimuli enclosed by triangles denote training exemplars for Category 2.)

As was the case in Experiment 1, two conditions were tested. In Condition EF all training exemplars were presented with equal frequency, whereas in Condition HF19 Face 19 was presented five times as often as the other exemplars. The purpose of the frequency manipulation was to obtain additional evidence bearing on the interactive roles of similarity and frequency in determining classification and recognition.

Method

Subjects

The subjects were 160 undergraduates from Indiana University who participated as part of an introductory psychology course requirement. There were 80 subjects each in Conditions EF and HF19.

Stimuli and Apparatus

The stimuli and apparatus were the same as in Experiment 1.

Procedure

The procedure was the same as in Experiment 1, except that the new category structure illustrated in Figure 9 was tested. In Condition EF all exemplars were presented with equal frequency during training, whereas in Condition HF19 Face 19 was presented five times as often as any of the other exemplars.

Results and Theoretical Analyses

Condition EF

The probability with which each stimulus was classified in Category 1 and was judged as "old" in Condition EF is presented in Table 5. As was the case in Experiment 1, there was a low correlation between the classification confidence measure and the probability of "old" recognition responses ($r = .22$). Various examples exist in which classification confidence is low but recognition probability is high (e.g., Faces 16, 24, 25, and 34), and likewise in which classification confidence is high yet recognition probability is low (e.g., Faces 5, 6, and 21).

Despite the low classification–recognition correlation, the exemplar model achieves fairly good quantitative fits to both sets of data, although the fits are not as impressive as those obtained in Experiment 1. The predicted Category 1 and "old" recognition response probabilities are presented along with the observed probabilities in Table 5, and scatterplots of the observed against predicted classification and recognition probabilities are shown in Figure 10. The exemplar model accounts for 86.3% of the variance in the classification data and for 81.5% of the variance in the recognition data. By comparison, the prototype model accounted for only 77.2% of the variance in the classification data and 47.3% of the variance in the recognition data. An augmented version of the exemplar model, which yields an appreciably better fit to the classification data than the standard version, is discussed in Appendix D. Because none of the major conclusions are changed and the augmented model involves the addition of some post hoc assumptions, I prefer to focus on the standard model in the main text.

The maximum likelihood parameters and summary fits for the exemplar model as well as the prototype model are reported in Table 6. Again, the best fitting distributions of attention weights differ considerably for classification and recognition. As predicted, for classification subjects attended primarily to eye separation and mouth height. Indeed, they came close to theoretically optimizing their classification performance: With the best fitting weights, the predicted proportion of correct classification responses is .791, which is close to the optimal of .819. The best fitting distribution of classification weights is plotted against the optimal distribution in Figure 11. As can be seen, the correspondence is quite good.

Also as predicted, subjects increased their attention to the dimensions of eye height and nose length in making their recognition judgments (see Table 6). Impressively, whereas subjects devoted nearly half of their attentional resources to mouth height for the classification task, this dimension was given zero weight for recognition. The best fitting distribution of recognition weights is not the theoretically optimal distribution, as can be seen in Figure 11. Nevertheless, had subjects adopted the same distribution of weights for recognition as they had adopted for classification, the predicted hit minus false-alarm rate would have been only .086. With the distribution of weights that was actually used, the predicted hit minus false-alarm rate was .229. Thus, subjects improved their recognition performance considerably by shifting some attention to the new dimensions. (The theoretically optimal hit minus false-alarm rate was .283.)

Condition HF19

The probability with which each stimulus was classified in Category 1 and was judged as "old" in Condition HF19 is reported in Table 5. To facilitate comparisons, Figure 12 plots these classification and recognition probabilities against those observed in Condition EF. The figure also highlights with solid circles the critical stimuli for which the exemplar model predicts the largest increases in classification and recognition probabilities. The critical stimuli were selected on the basis of their weighted distance to the high-frequency item (Face 19). The weights used in calculating the distances were the maximum likelihood parameters estimated in Condition EF. For classification, six faces were proximal (weighted distance less than .30) to Face 19 (Faces 11, 16, 19, 25, 26, and 29), whereas for recognition, five faces were proximal (Faces 13, 19, 22, 25, and 29).

Difference-score analyses analogous to those already described in Experiment 1 were conducted. For classification the frequency effect was robust: The average difference score for the 6 critical faces was .195, and for the remaining 28 noncritical faces it was −.030, $t(32) = 4.81$, $p < .01$. Indeed, the effect of the frequency manipulation extended even to the six faces that were next closest to the high-frequency face, which are represented by the cross-hatched circles in Figure 12A. As can be seen, there is a strong tendency for all of the solid and cross-hatched circles to lie above the diagonal in Figure 12A, whereas the open circles corresponding to nonproximal stimuli lie close to the diagonal or below it.

For recognition the average difference score for the 5 critical faces was .096 and for the remaining 29 noncritical faces it was −.026, $t(32) = 3.22$, $p < .01$. This pattern of results is similar to the one observed in Experiment 1. Interactive effects of frequency and similarity were observed for both the clas-

Table 5
Observed and Predicted Category 1 Response Probabilities and "Old" Recognition Probabilities for Each Face in Experiment 2

	Condition EF		Condition HF19			Condition EF		Condition HF19	
Face	$P(C_1)$	P(old)	$P(C_1)$	P(old)	Face	$P(C_1)$	P(old)	$P(C_1)$	P(old)
1	.469	.195	.549	.163	18	.366	.541	.250	.520
	.660	.126	.748	.151		.381	.563	.189	.579
2	.304	.666	.159	.645	19	.734	.687	.969	.791
	.253	.639	.151	.761		.774	.604	.963	.788
3	.559	.295	.575	.234	20	.097	.666	.055	.644
	.456	.369	.430	.342		.095	.671	.069	.730
4	.614	.286	.604	.217	21	.837	.354	.831	.362
	.375	.263	.415	.082		.944	.394	.938	.294
5	.727	.304	.642	.230	22	.860	.742	.947	.827
	.843	.264	.698	.132		.925	.818	.969	.781
6	.178	.445	.133	.396	23	.093	.779	.052	.795
	.156	.406	.101	.327		.094	.838	.044	.775
7	.840	.539	.836	.504	24	.806	.757	.876	.776
	.855	.610	.881	.531		.633	.715	.811	.767
8	.128	.642	.085	.667	25	.419	.842	.606	.884
	.113	.556	.126	.635		.394	.600	.675	.731
9	.575	.491	.385	.479	26	.618	.345	.809	.337
	.717	.516	.516	.434		.744	.388	.855	.396
10	.619	.360	.425	.315	27	.796	.804	.869	.845
	.681	.325	.541	.264		.675	.781	.731	.794
11	.618	.345	.809	.337	28	.855	.772	.843	.767
	.711	.421	.881	.381		.844	.775	.844	.706
12	.654	.505	.509	.518	29	.725	.566	.960	.637
	.763	.581	.656	.594		.838	.638	.950	.781
13	.854	.552	.930	.601	30	.366	.541	.250	.520
	.937	.673	.956	.731		.318	.643	.222	.513
14	.417	.549	.559	.502	31	.281	.247	.193	.197
	.338	.519	.516	.434		.270	.164	.126	.164
15	.677	.564	.720	.522	32	.230	.778	.112	.797
	.525	.575	.660	.491		.283	.799	.169	.819
16	.422	.758	.605	.770	33	.567	.373	.574	.288
	.394	.588	.679	.742		.369	.331	.409	.308
17	.125	.462	.081	.474	34	.505	.603	.405	.567
	.081	.588	.031	.566		.488	.675	.264	.667

Note. Top entries in each row are predicted probabilities; bottom entries are observed probabilities.

sification and recognition data, although the effects were again larger for classification than for recognition. Regarding the recognition data, note that the largest effect occurred for the frequency-manipulated item itself, namely Face 19 (difference score = .19). The average increase in recognition probability for the four neighbors of Face 19 was .072. By comparison, the average increase in probability with which the five neighbors of Face 19 were classified in Category 1 was .196. Further evidence that the frequency manipulation had a differential effect for classification and recognition is provided in the ensuing theoretical analyses.

In fitting the frequency-sensitive exemplar model to the Condition HF19 data, it was found necessary to allow the exemplar-strength parameter for Face 19 to vary freely across classification and recognition. The predicted probabilities of Category 1 responses and of "old" recognition judgments are

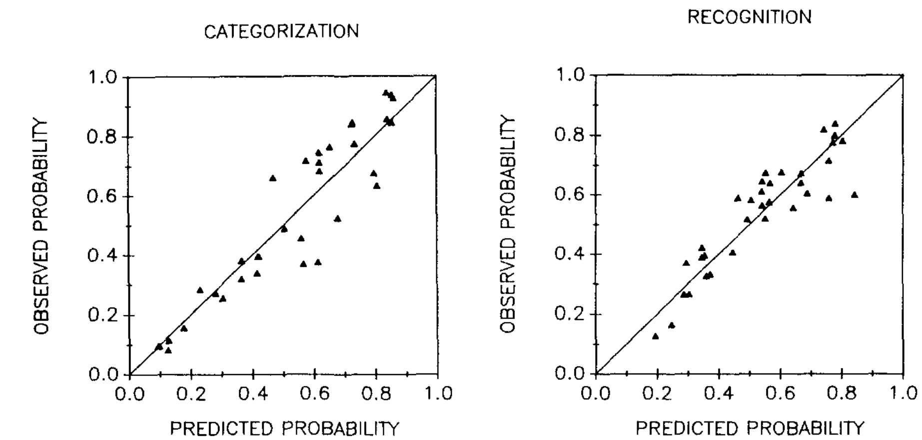

Figure 10. Scatterplot of observed against predicted classification and recognition probabilities in Condition EF of Experiment 2.

shown with the observed probabilities in Table 5. Scatterplots of the observed against predicted probabilities are presented in Figure 13. Although there is room for improvement, the fits are fairly good, with 91.9% and 89.0% of the variance in the classification and recognition data accounted for, respectively.

The maximum likelihood parameters and summary fits for the frequency-sensitive model are reported in Table 6. The distributions of attention weights for classification and recognition are similar to those derived in Condition EF. The most notable result concerns the estimates for the exemplar-strength parameter (M_{19}). For classification, the strength-parameter estimate was M_{19} = 2.73, whereas for recognition the M_{19} estimate did not depart from its lower bound of 1.0. (Slightly better fits to the recognition data can be achieved by allowing M_{19} to dip below 1.0, but the improvement is not statistically significant.) The parameters and summary fits for a restricted version of the exemplar model in which the M_{19} parameter was constrained to be constant across classification and recognition are shown in Table 6. The decrement in fit for this restricted model in relation to the full model with separate exemplar-strength parameters was statistically significant, $\chi^2(1, N = 10{,}880) = 48.4$, $p < .01$. This result should be interpreted with caution, however, because Table 6 indicates that there was little difference between the full and restricted models when fit is measured in terms of the sum of squared deviations between predicted and observed probabilities.

The exemplar model was also fitted to the Conditions EF and HF19 classification and recognition data simultaneously (see Table 7). In fitting the model to the classification data, all parameters were held constant across Conditions EF and HF19 except for the exemplar-strength parameter and the category response-bias parameter. The maximum likelihood estimate of the strength parameter in Condition HF19 was M_{19} = 3.18. The fit for a version of the model in which M_{19}

Table 6
Maximum Likelihood Parameters and Summary Fits for the Context Model and Prototype Model in Experiment 2

Model	Parameters									Fits		
	σ	c	w_1	w_2	w_3	w_4	x_c	b	M_{19}	SSE	% Var	-ln L
Condition EF												
Context												
Classification	.389[a]	1.169[a]	.108	.463	.000	.429		−.071		.340	86.3	219.1
Recognition	.389[a]	1.169[a]	.294	.499	.207	.000	4.748			.212	81.5	174.4
Prototype												
Classification	.044[a]	.092[a]	.128	.478	.011	.382		.003		.567	77.2	303.7
Recognition	.044[a]	.092[a]	.245	.295	.337	.123	1.902			.603	47.3	309.9
Condition HF19												
Context												
Classification	.406[a]	1.865[a]	.126	.587	.000	.288		.478	2.73	.277	91.9	194.8
Recognition	.406[a]	1.865[a]	.414	.378	.208	.000	3.187		1.00	.187	89.0	168.9
Restricted context												
Classification	.379[a]	1.719[a]	.142	.660	.000	.198		.113	1.54[a]	.296	91.4	205.2
Recognition	.379[a]	1.719[a]	.290	.529	.180	.000	3.719		1.54[a]	.211	87.6	182.6

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood.
[a] These parameters were held fixed across classification and recognition.

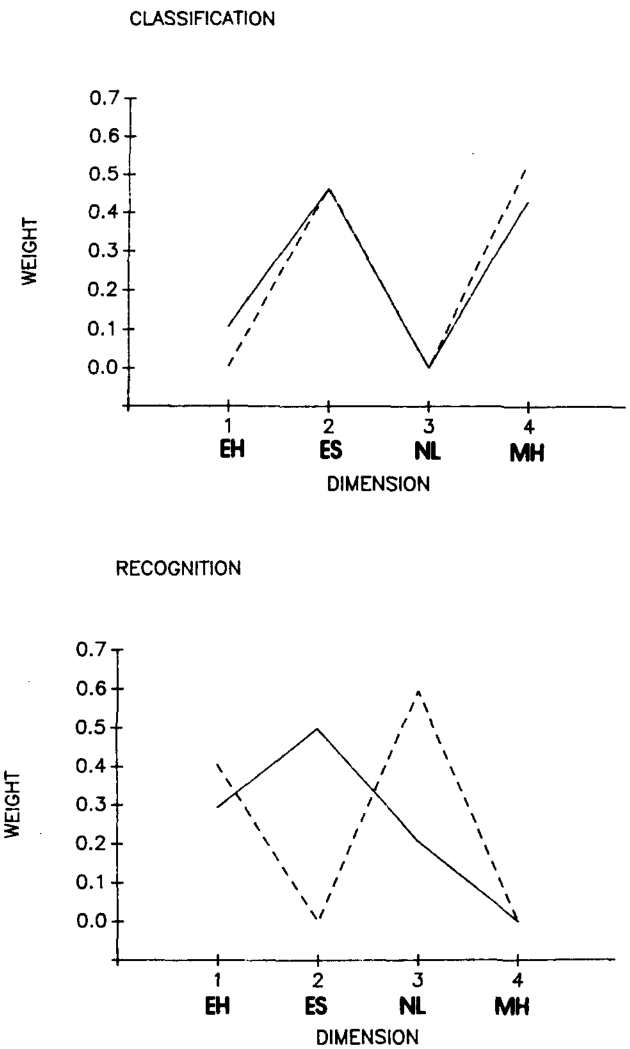

Figure 11. Best fitting distributions of attention weights (solid lines) plotted with ideal-observer distributions of attention weights (dashed lines) in Condition EF of Experiment 2.

was held fixed at 1.0 was dramatically worse than that of the full model, $\chi^2(1, N = 10{,}880) = 167.1$, $p < .01$. In fitting the model to the recognition data, all parameters were held constant across Conditions EF and HF19 except for the strength parameter and the recognition criterion. The maximum likelihood estimate of M_{19} in Condition HF19 was 1.40, which leads to a significantly better fit to the recognition data than if the M_{19} parameter is held fixed at 1.0, $\chi^2(1, N = 10{,}880) = 7.12$, $p < .05$. However, the magnitude of the best fitting exemplar-strength parameter for recognition, as well as the magnitude of the improvement in fit, seems small in relation to what was observed for the classification data.

To summarize, the theoretical analyses indicate that the frequency manipulation exerted a major impact on subjects' classification responses, and this impact appears to be interpretable in terms of changes in category response bias and exemplar strength across Conditions EF and HF19. Simple changes in bias and exemplar strength may not tell the whole story, however, because both the qualitative and model-based analyses suggest that the frequency manipulation had a weaker effect on recognition than on classification. This pattern of results is similar to the one observed previously in Experiment 1 and may indicate shortcomings in the present theoretical account of relations between classification and recognition.

General Discussion

The goal of this research was to provide rigorous quantitative tests of an exemplar-similarity model for relating perceptual classification and recognition memory. On the basis of similarity-ratings data, an MDS solution was derived for 34 schematic faces. This MDS solution was then used in conjunction with the exemplar model to predict classification and recognition performance. It was assumed that classification judgments were based on the similarity of a probe to the exemplars of a target category in relation to exemplars of a contrast category, whereas recognition judgments were based on absolute summed similarity to all stored examplars. Because classification and recognition involve different decision rules, performance in the two tasks may often be lowly correlated, as was observed in the present study. Despite the low correlations, fairly good quantitative accounts of the recognition and classification data were achieved in the present research in terms of the unifying assumption that both types of judgments are based on similarity comparisons with stored exemplars. The present demonstrations go beyond earlier ones reported by Nosofsky (1988a) by showing that detailed quantitative predictions of classification and recognition performance for individual items can be achieved within this proposed theoretical framework.

The model-based analyses also pointed toward the importance of selective attention processes in determining the classification–recognition relation. In previous work concerned with relations between identification and classification performance, Nosofsky (1986, 1987, 1989) observed that subjects' attentional weightings of the component dimensions of the stimuli shifted systematically with the structure of the categories to be learned. The present work shows that attentional resources may be allocated differentially for recognition judgments as well. Indeed, evidence was provided in the present study that the distributions of attention over the psychological dimensions differed depending on whether subjects made classification or recognition judgments. The utility of attending to particular dimensions varies with task goals. A dimension that is highly diagnostic for discriminating between categories may be useless for discriminating between old and new items. The present study provided support for the hypothesis that the learners distributed attention over the psychological dimensions to optimize their classification performance (cf. Nosofsky, 1984, 1986; Reed, 1972; Shepard et al., 1961), and some tendencies in that direction were observed for recognition. Note that recognition judgments were not made until after all training exemplars had been presented. Thus, the attention weights must be reflecting, at least in part, processes operating at the time of retrieval and decision making rather than solely at time of storage of the exemplar information. An important issue for future work involves the development of process models that specify

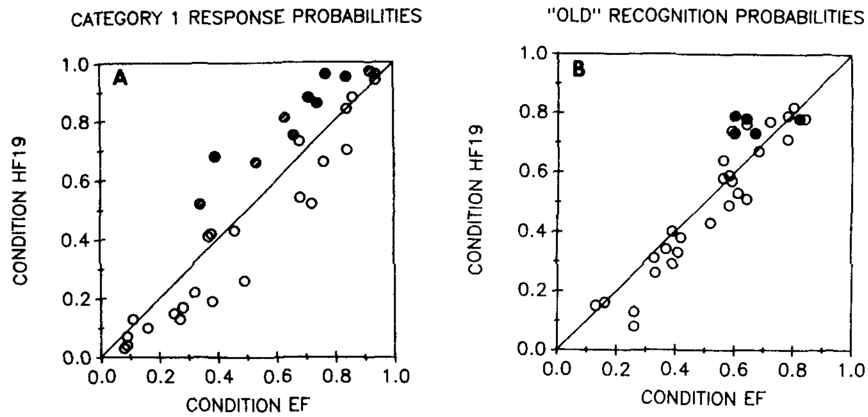

Figure 12. Observed Category 1 response probabilities in Condition HF19 plotted against observed Category 1 response probabilities in Condition EF (Panel A), and observed "old" recognition probabilities in Condition HF19 plotted against observed "old" recognition probabilities in Condition EF (Panel B). (In both panels, the results for the critical stimuli [the high-frequency item and its neighbors] are marked by the solid circles, and the results for the noncritical stimuli are marked by the open circles. The cross-hatched circles mark results for stimuli that are proximal to the high-frequency item but not as proximal as the critical stimuli.)

mechanisms of weight change as a function of classification and recognition experience.

The present research also pointed toward the importance of individual item frequency in classification and recognition. Systematic changes in classification and recognition probabilities were observed as a function of a frequency manipulation for an individual item. These changes were well predicted by a model which assumed that the exemplar memory strength associated with an item increased as its presentation frequency was increased. Because exemplar strength was assumed to combine multiplicatively with interitem similarity, the model predicted an interactive effect of frequency and similarity, and this interactive effect was observed.

Perhaps the major shortcoming of the present theoretical account, however, involves the complex effect that frequency and similarity exerted on recognition performance, and in particular the reason that the frequency effects were weaker for recognition than for classification in both Experiments 1 and 2. A post hoc explanation is the following. Various researchers have suggested that perceptual differentiation may result from increased experience with objects (e.g., Gibson & Gibson, 1955). Indeed, in model-based analyses of identification confusion data, Nosofsky (1987) provided evidence of decreases in similarities among objects as a function of learning. To the extent that increased perceptual differentiation occurred in the local region of the psychological space around the high-frequency items, recognition probabilities for highly similar distractor items would tend to be lowered. Thus, with regard to recognition, the frequency manipulation may have exerted competing influences: increased exemplar strength for the high-frequency items but decreasing perceptual similarity between the strengthened items and their neighbors in the

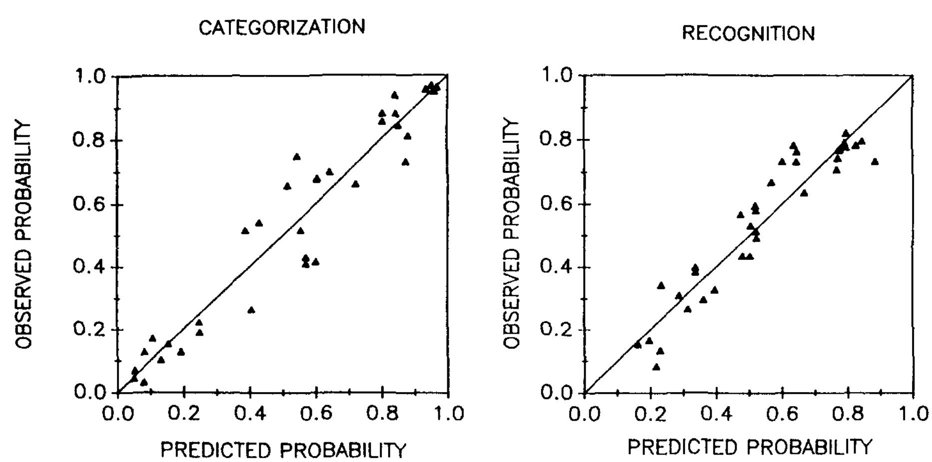

Figure 13. Scatterplot of observed against predicted classification and recognition probabilities in Condition HF19 of Experiment 2.

Table 7
Maximum Likelihood Parameters and Summary Fits for the Context Model When the Conditions EF and HF19 Classification and Recognition Data Are Fitted Conjointly

	Parameters								Fits		
Condition	σ	c	w_1	w_2	w_3	w_4	b	M_{19}	SSE	% Var	-ln L
Classification											
EF	.341[a]	1.095[a]	.127[a]	.479[a]	.000[a]	.394[a]	−.058	1.000	.338	86.4	217.0
HF19	.341[a]	1.095[a]	.127[a]	.479[a]	.000[a]	.394[a]	1.023	3.180	.308	91.0	201.9
Recognition											
EF	.464[a]	1.675[a]	.356[a]	.434[a]	.211[a]	.000[a]	3.511	1.000	.204	82.2	172.7
HF19	.464[a]	1.675[a]	.356[a]	.434[a]	.211[a]	.000[a]	3.677	1.404	.216	87.3	177.5

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood. The value of M_{19} was held fixed at 1.000 in Condition EF.
[a] These parameters were held fixed across Conditions EF and HF19.

psychological space (cf. Nosofsky, 1988b, 1988c; Shiffrin, Ratcliff, & Clark, 1990). Furthermore, it can be argued that this increasing perceptual differentiation would not weaken the exemplar-strength effect in the classification paradigm, because classification and recognition involve different task goals. In deciding category membership for a novel item that is similar to a strong item, an observer may be able to perceptually discriminate the objects. But from a cognitive, judgmental standpoint, the high similarity of the novel item to the strong training exemplar should still lead the observer to classify it in the strong exemplar's category. In other words, two forms of similarity may be involved: a "perceptual" similarity that places limits on an observer's ability to discriminate items, and a "cognitive" similarity that leads to generalizations in decisions about class membership (cf. Ennis, 1988; Estes, 1986a; Nosofsky, 1987; Shepard, 1986).

Another important question for future research concerns the generality of the summed-similarity rule for recognition. People may avail themselves of a number of alternative strategies in making recognition judgments. Indeed, various researchers have suggested that in addition to basing recognition judgments on a global familiarity index, people may make use of search and retrieval strategies (e.g., Atkinson & Juola, 1974; Mandler, 1980; Tulving & Thomson, 1971). The global familiarity rule may be prevalent in classification learning situations, where it is presumably difficult to gain unique access to memory representations of similar stimuli. But a complete account of the classification–recognition relation may require recourse to search and retrieval strategies as well, particularly if highly distinctive training items are used. Furthermore, a natural extension of the present investigation, already initiated by Clark (1988), is to study how people recall individual category exemplars in addition to how they classify and recognize them. And in the act of recall, search strategies are likely to be paramount.

References

Anderson, J. R., Kline, P. J., & Beasley, C. M. (1979). A general learning theory and its application to schema abstraction. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 277–318). New York: Academic Press.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 33–53.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 598–612.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol 1. Learning, memory, and thinking* (pp. 243–293). San Francisco: Freeman.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 638–648.

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 2* (pp. 57–105). San Francisco: Freeman.

Clark, S. E. (1988). *A theory for classification and memory retrieval.* Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.

Cox, J. A., & Wallsten, T. S. (1987). *Evidence for the use of configural properties in the classification of faces.* Unpublished manuscript.

Ennis, D. M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General, 117,* 408–411.

Ennis, D. M., Palen, J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology, 32,* 449–465.

Estes, W. K. (1986a). Array models for category learning. *Cognitive Psychology, 18,* 500–549.

Estes, W. K. (1986b). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General, 115,* 155–174.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Getty, D. J., Swets, J. B., Swets, J. A., & Green, D. M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics, 26,* 1–19.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for multiple-trace hypothesis. *Journal of Experimental Psychology, 88,* 297–306.

Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116–122.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Unpublished manuscript.

Lockhead, G. R. (1970). Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology, 85,* 1–10.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics, 20,* 49–54.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Medin, D. L. (1986). Comment on "Memory storage and retrieval processes in category learning." *Journal of Experimental Psychology: General, 115,* 373–381.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117,* 68–85.

Medin, D. L., & Reynolds, T. J. (1985). Cue-context interactions in discrimination, categorization, and memory. In P. D. Balsam & A. Tomie (Eds.), *Context and learning* (pp. 323–356). Hillsdale, NJ: Erlbaum.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

Metcalfe, J., & Fisher, R. P. (1986). The relation between recognition memory and classification learning. *Memory & Cognition, 14,* 164–173.

Metcalfe-Eich, J. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627–661.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Nosofsky, R. M. (1983). Shifts of attention in the identification and discrimination of intensity. *Perception & Psychophysics, 33,* 103–112.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–109.

Nosofsky, R. M. (1988a). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M. (1988b). On exemplar-based exemplar representations: Reply to Ennis (1988). *Journal of Experimental Psychology: General, 117,* 412–414.

Nosofsky, R. M. (1988c). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

Nosofsky, R. M. (1989). Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception & Psychophysics, 45,* 279–290.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 282–304.

Omohundro, J. (1981). Recognition vs. classification of ill-defined category exemplars. *Memory & Cognition, 9,* 325–331.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics, 14,* 565–569.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304–308.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382–407.

Reed, S. K. (1988). *Cognition: Theory and applications.* Belmont, CA: Wadsworth.

Reed, S. K., & Friedman, M. P. (1973). Perceptual vs. conceptual categorization. *Memory & Cognition, 1,* 157–163.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology, 55,* 509–523.

Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General, 115,* 58–61.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75*(13, Whole No. 517).

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179–195.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory. *Journal of Experimental Psychology, 87,* 352–373.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology, 5,* 102–122.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology.* San Francisco: Freeman.

Zelen, M., & Severo, N. C. (1972). Probability functions. In M. Abramowitz & I. Stegun (Eds.), *Handbook of mathematical functions.* New York: Dover.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika, 48,* 27–48.

Appendix A

The physical specifications for the 34 schematic faces and the MDS solution for the faces that was derived from the similarity judgment data are provided in Table A1. With regard to the physical specifications, eye height is measured in millimeters from the top of the outline of the face; eye separation is measured in millimeters between the center of the two eyes; nose length is measured in millimeters; and mouth height is measured in millimeters from the bottom of the outline of the face.

Table A1
Physical Specifications and Multidimensional Scaling (MDS) Coordinates for the 34 Faces

	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
Face	EH	MDS	ES	MDS	NL	MDS	MH	MDS
1	23.5	−1.025	21.5	0.493	13.5	0.048	16.5	−0.666
2	19.5	−0.172	11.5	−0.557	18.0	0.337	12.0	0.163
3	23.5	−0.980	16.5	0.275	13.5	−0.005	12.0	−0.067
4	23.5	−0.951	16.5	0.259	18.0	0.399	12.0	0.093
5	23.5	−0.960	16.5	0.198	18.0	0.380	7.5	0.527
6	15.0	0.665	11.5	−0.441	9.0	−0.508	16.5	−0.396
7	19.5	−0.059	16.5	0.243	9.0	−0.602	7.5	0.624
8	15.0	0.586	11.5	−0.511	18.0	0.381	16.5	−0.507
9	15.0	0.823	11.5	−0.539	18.0	0.332	7.5	0.633
10	15.0	0.823	11.5	−0.504	9.0	−0.487	7.5	0.776
11	19.5	−0.227	21.5	0.589	9.0	−0.529	16.5	−0.431
12	23.5	−1.041	11.5	−0.538	9.0	−0.449	7.5	0.629
13	15.0	0.788	21.5	0.597	13.5	0.062	12.0	0.175
14	19.5	−0.199	16.5	0.186	9.0	−0.572	16.5	−0.429
15	19.5	−0.106	16.5	0.150	9.0	−0.604	12.0	0.129
16	15.0	0.622	16.5	0.192	13.5	0.003	16.5	−0.497
17	23.5	−1.158	11.5	−0.526	9.0	−0.371	16.5	−0.530
18	15.0	0.765	11.5	−0.491	13.5	0.086	12.0	0.083
19	15.0	0.553	21.5	0.623	18.0	0.461	16.5	−0.412
20	19.5	−0.275	11.5	−0.578	18.0	0.295	16.5	−0.471
21	19.5	0.018	21.5	0.617	9.0	−0.586	7.5	0.621
22	15.0	0.665	21.5	0.507	18.0	0.476	12.0	0.116
23	19.5	−0.207	11.5	−0.546	13.5	−0.064	16.5	−0.488
24	15.0	0.683	16.5	0.215	13.5	0.069	12.0	0.167
25	15.0	0.495	16.5	0.218	18.0	0.445	16.5	−0.481
26	19.5	−0.227	21.5	0.589	9.0	−0.529	16.5	−0.431
27	15.0	0.723	16.5	0.188	18.0	0.362	12.0	0.142
28	19.5	−0.020	16.5	0.211	13.5	−0.095	7.5	0.723
29	15.0	0.591	21.5	0.645	13.5	0.025	16.5	−0.438
30	15.0	0.765	11.5	−0.491	13.5	0.086	12.0	0.083
31	23.5	−1.114	11.5	−0.520	18.0	0.636	12.0	−0.028
32	19.5	−0.154	11.5	−0.562	13.5	−0.043	12.0	0.057
33	22.7	−0.856	16.5	0.197	16.2	0.241	12.0	0.007
34	15.9	0.704	12.5	−0.287	12.6	−0.164	11.1	0.178

Note. EH = eye height, ES = eye separation, NL = nose length, and MH = mouth height.

Appendix B

Given the assumptions stated in the text, the random variable $E_{1,i} - E_{2,i}$ is normally distributed with mean $\sum_{j \in C_1} s_{ij} - \sum_{j \in C_2} s_{ij}$ and variance $10\sigma^2$. The probability of making a Category 1 response given presentation of stimulus i is found, therefore, by integrating this normal random variable from b to ∞, or alternatively, integrating a standardized normal random variable from $[\sum_{j \in C_2} s_{ij} - \sum_{j \in C_1} s_{ij} + b]/\sqrt{10}\sigma$ to ∞. Because of the symmetry of the standardized normal, this is equivalent to computing the integral from $-\infty$ to $[\sum_{j \in C_1} s_{ij} - \sum_{j \in C_2} s_{ij} - b]/\sqrt{10}\sigma$ of a standardized normal random variable. Analogously, the theoretical expression for the probability of an "old" recognition response is found by computing the integral from $-\infty$ to $[\sum_{j \in C_1} s_{ij} + \sum_{j \in C_2} s_{ij} - x_c]/\sqrt{10}\sigma$ of a standardized normal random variable.

A numerical approximation (accurate to within 7.5×10^{-8}) for computing $P(x) = \int_{-\infty}^{x} Z(t)dt$, where $Z(t)$ is the standardized normal random variable [i.e., $Z(t) = 1/\sqrt{2\pi}\, e^{-t^2/2}$], and where $x \geq 0$, is the following (from Zelen & Severo, 1972, p. 932):

$$P(x) \approx 1 - Z(x)(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5),$$

where $t = (1 + px)^{-1}$, $p = .2316419$, $b_1 = 0.319381530$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, and $b_5 = 1.330274429$.

Appendix C

The psychological prototypes are computed by averaging over the MDS coordinate values associated with the five exemplars of each category. Thus, the coordinates for the Category 1 prototype are $p_{11} = -.818$, $p_{12} = .134$, $p_{13} = .240$, and $p_{14} = .010$; the coordinates for the Category 2 prototype are $p_{21} = .568$, $p_{22} = -.350$, $p_{23} = -.177$, and $p_{24} = .226$. The distance between stimulus i and prototype j is computed by using a (weighted) Euclidean metric:

$$d_{ij} = c\left[\sum_{m=1}^{4} w_m \left|x_{im} - p_{jm}\right|^2\right]^{1/2},$$

where the parameters are defined as before. This distance is transformed to a similarity measure by using an exponential decay function, $s_{ij} = \exp(-d_{ij})$. The degree to which stimulus i activates prototype j is given by $a_{ij} = s_{ij} + e_j$, where the e_js are independent and identically distributed normal random variables with mean 0 and variance σ^2. The evidence for Category 1 is equal to the degree to which Prototype 1 is activated, and likewise for the evidence for Category 2. The decision rules for classification and recognition are the same as for the exemplar model (Inequalities 5 and 6). The predicted classification and recognition probabilities are computed as explained in Appendix B, except the decision variables have variance $2\sigma^2$, rather than $10\sigma^2$.

Appendix D

In exploratory analyses, it was discovered that a substantially better fit to the Conditions EF and HF19 classification data could be achieved by incorporating an additional process into the formal modeling. A central theme in previous tests of the exemplar model and of the present work has been that similarities among exemplars are modifiable by selective attention. The attention process has been formalized by using weight parameters in the distance function (Equation 1). The geometric interpretation of the weights is one of stretching or shrinking of the psychological space along its coordinate axes. This formalization has assumed that the attention process acts globally on each of the psychological dimensions. However, with regard to modeling classification data, researchers such as Luce, Green, and Weber (1976), Medin and Edelson (1988), Nosofsky (1983), and Nosofsky, Clark, and Shin (1989) have argued for the potential importance of value-specific selective attention. By this I mean devoting attention to a particular value or local region of a psychological dimension, as opposed to attending globally to the entire dimension. With reference to Figure 9A, note that the very large values of eye separation and mouth height are sufficient cues for membership in Category 2. It was hypothesized that subjects might have noticed this relation and selectively attended to these values when learning the classification.

To formalize the value-specific selective attention process, it is assumed that there is local stretching of the psychological space in the region of the attended values. Specifically, the eye separation coordinates for all stimuli having the large values of eye separation were incremented by a free parameter v, as were the mouth-height coordinates for all stimuli having the large values of mouth height. The maximum likelihood parameters and summary fits for this augmented version of the exemplar model are reported in Table D1. Adding the value-specific selective attention parameter dramatically improved the fit of the exemplar model to the classification data of both Conditions EF and HF19. Scatterplots of the observed against predicted probabilities for the augmented model are shown in Figure D1. Comparison with Figures 10A and 13A reveals a noticeable improvement. The remaining parameters have values that were similar to ones estimated previously when fitting the standard model, so the earlier conclusions regarding the patterns of global selective attention and the role of frequency remain unchanged. This exploratory analysis should be regarded as only suggestive, and further research is clearly needed to provide systematic evidence for the operation of value-specific selective attention processes.

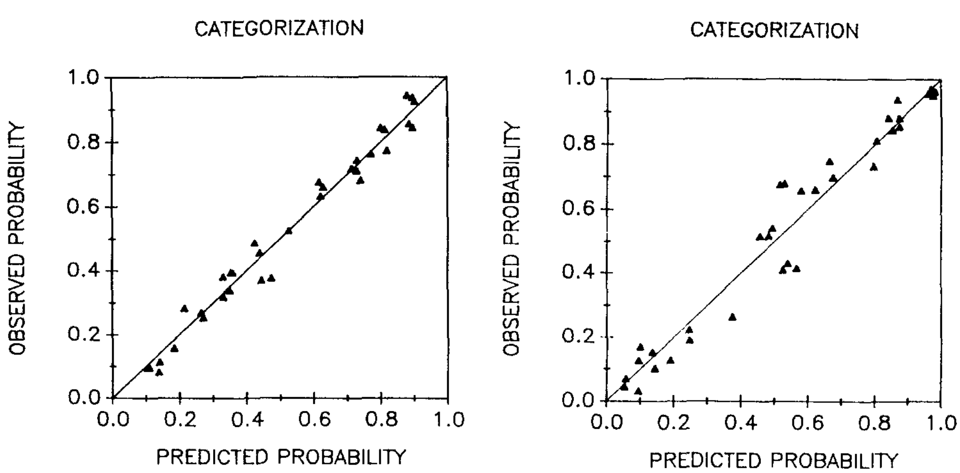

Figure D1. Scatterplot of observed against predicted classification probabilities in Conditions EF and HF19 of Experiment 2 for the exemplar model with value-specific selective attention.

Table D1
Maximum Likelihood Parameters and Summary Fits for the Context Model With Value-Specific Selective Attention

Condition EF	Parameters										Fits		
	σ	c	W_1	W_2	W_3	W_4	X_c	b	v	M_{19}	SSE	% Var	-ln L
EF													
Classification	.437[a]	1.333[a]	.132	.439	.000	.429		−.239	.247		.059	97.6	115.2
Recognition	.437[a]	1.333[a]	.313	.481	.206	.000	4.291				.201	82.5	169.9
HF19													
Classification	.369[a]	1.395[a]	.187	.582	.002	.229		.206	.311	2.536	.165	95.2	151.0
Recognition	.369[a]	1.395[a]	.400	.332	.268	.000	4.047			.781	.211	87.6	175.0

Note. SSE = sum of squared deviations between predicted and observed probabilities; % Var = percentage of variance accounted for; ln L = log-likelihood.
[a] These parameters were held fixed across classification and recognition.

Received September 19, 1989
Revision received April 9, 1990
Accepted April 10, 1990 ■